Structural Studies on Protein Disulphide Isomerase

by

Ateesh Sidhu

A thesis submitted in partial fulfilment

of the requirements for the degree of

Doctor of Philosophy

The University of Warwick

Department of Biological Sciences

April 2008

DECLARATION

I hereby declare that the research submitted in this thesis was conducted by myself under the supervision of Prof. R. B. Freedman at the Department of Biological Sciences, University of Warwick.

No part of this work has previously been submitted to be considered for a degree or other qualification. All sources of information have been specifically acknowledged in the form of references.

Ateesh Sidhu

April 2008

ACKNOWLEDGEMENTS

I would like to begin by thanking my supervisor Professor Robert Freedman, for giving me this opportunity to study such a challenging and interesting project and for providing me with constant support and encouragement throughout this project. I am indebted to Dr. K. Wallis for all her guidance and patience in the laboratory, without her expertise this project would have been a great struggle.

I would also like to thank Dr. R. Williamson, Dr. M. Howard and Dr. L. Byrne (University of Kent) for all their advice and assistance with the NMR experiments and data analysis. I am grateful for all your patience and time during my visits to Canterbury, which were always very productive and left me motivated to do more.

I would like to thank Dr. L. Ruddock, without your constructs and foundation work this project would not have been possible. I was also very privileged to have the opportunity to visit Prof. C.C. Wang's laboratory (Institute of Biophysics, Beijing, China). I would like to extend my gratitude to Prof. CC. Wang and Lei Wang, for their hospitality and making my time in Beijing experimentally productive.

My thanks also go to all the members past and present of the Freedman and Structural Biology group who made working in the laboratory so enjoyable. With a special thanks to Dr. G. de Pascale and Tom Clarke who made the long arduous hours and days of writing this thesis bearable.

Lastly and most importantly, I wish to thank the most important people in my life, my family and Emily, you have all supported me, encouraged me and generally put up with me, for that I am eternally grateful. Without your love and support I would not be the man I am and made it through this PhD. To them I dedicate this thesis.

ABSTRACT

Protein disulphide isomerase (PDI; EC 5.3.4.1) is a multifunctional enzyme which resides in the lumen of the endoplasmic reticulum (ER). It is approximated that over one-third of all human proteins fold in the ER. PDI is one of the main folding catalysts, specifically facilitating native disulphide bond formation.

PDI has four domains, three of which have been solved and all possessing a thioredoxin-type fold consisting of the β-α-β-α-β-α-β-β-α structure. Due to the intransigent nature of the **b'** domain of PDI, it remains structurally un-elucidated. The **b'** domain is vitally important as it holds the principal binding site and is essential for PDI activity.

The investigation of the two species which were produced during expression of most **b'** domain containing constructs led to the further biophysical analysis and identification of the monomer and dimer species. It was found that fractionation of the monomer and dimer species was vital in obtaining high resolution NMR data.

An elaborate method using assignments from the **b** domain and **b'x** were used, and proved very effective in achieving the goal of full backbone assignment of the large 27.5 kDa **bb'x** protein molecule. The fruit of this labour is that it allowed further probing dynamic and more detailed molecular NMR analysis.

NMR analysis identified evidence that the **bb'x** monomer species can exist in two forms, the closed form where the **x**-region binds to the **b'** domain and open form where the **x**-region is unbound. Chemical shift analysis revealed several key residues involved in protein flexibility and chemical shift mapping revealed the interface between the **b** domain and **b'x**. Defining the interface described here offers a method to model the full length PDI domain structural organisation; which can be used to more clearly define PDI function and how the domains of PDI are coordinated in protein function.

More detailed NMR analysis of relaxation dynamics (T_1 and T_2) revealed significantly differing motions attributed to the **b**, **b'** domains and **x**-region in the **bb'x** construct. The slower **b'** domain motion is believed to be related to the substrate binding function, where the **b'** domain appears to be gently flexing, in search for a substrate molecule.

CONTENTS

Declaration i

Acknowledgements ii

Abstract iii

Contents iv

List of Figures x

List of Tables xviii

Abbreviations xix

Chapter 1. General Introduction 1

1.1. Introduction 1

1.2. Protein Disulphide Isomerase 6

1.2.1. Activity 6

1.2.2. Abundance 6

1.2.3. PDI location 7

1.2.4. Redox-Isomerase function 7

1.2.4.1. Ero1 8

1.2.4.2. Oxidised glutathione 10

1.2.4.3. Flavin-dependent sulfhydryl oxidase 11

1.2.5. Thioredoxin fold 12

1.2.6. PDI structure 14

1.2.6.1. a and a' domains 15

1.2.6.2. C-terminal extension 17

1.2.6.3. b and b' domain 17

1.2.6.4. x-region 19

1.2.7. PDI function 20

1.2.8. Molecular chaperone function 22

1.2.9. PDI as a subunit of collagen prolyl 4-hydroxylase 23

1.2.10. PDI and MTP 26

1.3. PDI Family 28

1.3.1. Human PDI family members and structurally determined related proteins 28

1.3.2. Structural/functional data available for PDI family and related protein 31

1.3.3. ERp57....32
1.3.4. Calsequestrin....34
1.3.5. Yeast – PDI1p....35
1.3.6. Thermophilic fungus b' (2djk)....37
1.3.7. Human bb' chemical shifts....38
1.4. Protein NMR Spectroscopy....39
1.4.1. Basic principles of NMR....41
1.4.2. Chemical shift....43
1.4.3. HSQC....44
1.4.4. Through-bond experiments for sequential assignment....45
1.4.4.1. HNCACB....45
1.4.4.2. CBCA(CO)NH....46
1.4.4.3. TROSY....46
1.4.4.4. HNCA....47
1.4.4.5. TROSY - HN(CO)CA....47
1.4.5. Through-space experiments....48
1.4.5.1. T_1 relaxation....49
1.4.5.2. T_2 relaxation....50
1.4.5.3. HetNOE....51
1.4.6. Deuteration....51
1.4.7. Secondary structure prediction....51
1.4.8. Chemical shift mapping....52
1.4.9. Hydrogen/deuterium exchange....53
1.5. Aims of this Study....53
Chapter 2. General Materials and Methods....56
2.1. Materials....56
2.2. Protein Expression and Analytical methods....56
2.2.1. Expression constructs....57
2.2.2. Preparation of E.*coli* competent cells for DNA transformation....59
2.2.3. DNA transformation of E. *coli* cells....59
2.2.4. Preparation of plasmid DNA....60
2.2.5. DNA sequencing of plasmid constructs....60
2.2.6. Preparation of E. *coli* glycerol stocks....60
2.2.7. Expression of recombinant protein in *E.coli*....61

2.2.8. Determination of protein concentration....62
2.2.9. LB broth and agar....62
2.2.10. Minimal growth medium....62
2.2.11. SDS-polyacrylamide gel electrophoresis....63
2.2.12. PAGE under native conditions....65
2.2.13. Reduction and carboxymethylation....65
2.3. Chromatography Methods....66
2.3.1. Affinity chromatography....66
2.3.2. Anion-exchange chromatography....68
2.3.3. Size exclusion chromatography....68
2.4. Hydrodynamic methods....69
2.4.1. Analytical ultra-centrifugation – sedimentation velocity experiments....69
2.5. Spectroscopic Methods....70
2.5.1. Electrospray mass spectrometry....70
2.5.2. Circular dichroism spectroscopy....70
2.5.3. Nuclear Magnetic Resonance Spectroscopy....72
2.5.3.1. Preparation of protein samples for NMR....72
2.5.3.2. Data acquisition and processing....72
2.5.3.3. $^{15}N/^{1}H$ HSQC....73
2.5.3.4. Triple resonance experiments....73
2.5.3.5. Sequential backbone assignment....74
2.5.3.6. Secondary structure prediction using TALOS....76
2.5.3.7. Chemical shift mapping....76
2.5.3.8. T_1 and T_2 relaxation experiments....77
2.5.3.9. Heteronuclear NOE experiments....77
2.5.3.10. Hydrogen/deuterium exchange experiments....78
2.6. Structure Modelling....78
2.6.1. Molecular modelling....78
Chapter 3. Preparation and Preliminary Characterisation of PDI Proteins....79
3.1. Introduction....79
3.2. b domain....80
3.2.1. Expression and characterisation of the b domain....80

3.2.2. Optimisation of b domain expression.....84
3.2.3. Expression and purification of ^{15}N-labelled b domain 87
3.3. bb'x domain.....91
3.3.1. Expression and characterisation of the bb'x domain 92
3.3.2. Optimisation of bb'x expression.....102
3.3.3. Expression and purification of ^{15}N bb'x.....106
3.3.4. Expression and purification of ^{15}N/^{13}C bb'x.....113
3.4. b'x domain.....116
3.4.1. Expression and purification of ^{15}N b'x.....116
3.4.2. Expression and purification of ^{13}C/^{15}N b'x.....122
3.5. bb' domain 125
3.5.1. Expression and characterisation of the bb' domain 125
3.6. b' domain 129
3.6.1. Expression and characterisation of the b' domain 129
3.7. Full Length PDI 134
3.7.1. Expression and characterisation of full length PDI 134
3.8. Characterisation of the Pool 1 and Pool 2 species identified during purification.....138
3.8.1. Summary of observations during purification.....138
3.8.2. Analysis of storage of fractionated samples 140
3.8.3. Further analysis of the monomer/dimer species by AUC.....142
3.8.4. Characterisation of the high molecular weight species or dimer by SDS and native-PAGE.....143
3.9. Discussion.....147
Chapter 4. Preparation of ^{2}H/^{13}C/^{15}N-labelled bb'x 149
4.1. Introduction 149
4.2. Adaptation of Cells and Expression of Protein in Increasing D_2O Conditions 150
4.2.1. Test expression of adapted cells 151
4.3. Expression of ^{2}H Labelled Protein Using Adapted Cells and a Shock Expression Technique.....154
4.3.1. Test expression using the cells adapted to 86% and 96% D_20 conditions 154

4.3.2. Production of a mass spectrometry sample to assess incorporation efficiency 157
4.3.3. Test expression using un-adapted cells and shock expression to test the necessity of adaptation 160
4.4. Use of Isotopically Labelled Rich Medium with Adapted Cells to Increase Isotope Incorporation Efficiency 162
4.4.1. Test expression using 5% and 10% Celtone ^{2}H rich medium ... 163
4.4.2. Test expression using cells adapted to 96% D_2O grown in Spectra9 rich medium and shock induced 167
4.5. Production of a $^{2}H/^{13}C/^{15}N$ Sample for NMR Experiments 171
4.6. Discussion 177
Chapter 5. Initial NMR Studies on PDI Domain Combinations 178
5.1. Introduction 178
5.2. Prior NMR Studies on PDI Domain Combinations 179
5.2.1. b' domain 179
5.2.2. bb' domain 181
5.2.3. bb'x domain 182
5.2.4. b'x domain 183
5.3. Initial NMR Data on New PDI Domain Combination Samples 185
5.3.1. bb'x domain 185
5.3.2. b'x domain 188
5.3.3. b domain 189
5.4. Discussion 190
Chapter 6. NMR Backbone Assignment of bb'x 191
6.1. Introduction 191
6.2. First Attempt at bb'x Backbone Assignment and Stability Analysis 192
6.2.1. Backbone assignment of bb'x at 25 °C 192
6.2.2. Stability experiment at 35°C 194
6.2.3. Stability experiment at 40°C 195
6.3. Backbone Assignment using 40°C Triple Resonance Data 197
6.3.1. Initial backbone assignment attempt 198
6.3.2. b domain cross assignment 202

6.3.3. b'x backbone assignment to assist the assignment of the bb'x backbone 203
6.3.4. Use of a $^{2}H/^{13}C/^{15}N$ sample and TROSY experiments to complete assignments of bb'x 203
6.4. Discussion.....208
Chapter 7. Protein Conformation and Dynamics.....210
7.1. Introduction 210
7.2. Alternate Conformations 210
7.3. Chemical Shift Analysis.....212
7.3.1. bb'x chemical shifts associated with temperature.....213
7.3.2. Chemical shift mapping 216
7.3.3. Secondary structure prediction: TALOS 220
7.4. Hydrogen/Deuterium Exchange 222
7.4.1. b domain H/D exchange 222
7.4.2. bb'x H/D exchange 225
7.5. Dynamic Analysis using Relaxation Experiments 227
7.5.1. Heteronuclear-NOE.....227
7.5.2. T_1 and T_2 relaxation data 229
7.5.3. T_1 and T_2 rates to determine isotropy or anisotropy.....232
7.6. Discussion.....233
Chapter 8. Discussion.....235
8.1. Introduction 235
8.2. Expression and Characterisation of PDI Proteins 236
8.3. NMR Approach to Obtain b' Backbone Assignments.....237
8.4. Protein Conformation and Dynamics 241
References 244
Appendix 1 – Manuscript in preparation 264
Appendix 2 - Chemical shift data for bb'x at 25°C. 284
Appendix 3 - Chemical shift data for b'x at 25°C. 289
Appendix 4 - H/D exchange experimental data 292
Appendix 5 – Nucleotide and protein sequences 297

LIST OF FIGURES

Figure 1-1. Schematic diagram of protein folding and degradation in the endoplasmic reticulum. 5

Figure 1-2. Yeast Ero1 structure, taken from Sevier and Kaiser, 2000b. 8

Figure 1-3. A Graphical representation of the proposed electron flow, disulphide bond formation pathway for the oxidation and isomerisation of protein substrates. 10

Figure 1-4. Structural diagram of Human thioredoxin. 13

Figure 1-5. PDI domain architecture.I. 15

Figure 1-6. Structural diagram of the human PDI a domain. 16

Figure 1-7. Diagram of the human PDI b domain. 18

Figure 1-8. Reaction scheme for disulphide oxidation, reduction and isomerisation by PDI. 20

Figure 1-9. Model of human PDI structure using SAXS data. 32

Figure 1-10. Secondary structure diagram of b, b' domains and x-region at the C-terminus of ERp57. 34

Figure 1-11. Secondary structure diagram of the Calsequestrin. 35

Figure 1-12. Ribbon diagram of PDI1p domains.. 37

Figure 1-13. Ribbon diagram of Thermophilic fungus Humicola insolens b'. 38

Figure 1-14. Diagram representing the motion of a nucleus in an external magnetic field. 41

Figure 1-15. Diagram showing the energy levels for a nucleus. 42

Figure 1-16. Diagram showing the z axis magnetisation 43

Figure 1-17. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb'x domain at 40°C used to highlight the average chemical shifts for residues groups. 44

Figure 1-18. A dipeptide segment of a protein backbone showing correlations obtained in HNCACB experiments. 45

Figure 1-19. A dipeptide segment of a protein backbone showing correlations obtained in CBCA(CO)NH experiments. 46

Figure 1-20. A dipeptide segment of a protein backbone showing correlations obtained in HNCA experiments. 47

Figure 1-21. A dipeptide segment of a protein backbone showing correlations obtained in HN(CO)CA experiments. 48
Figure 1-22. Diagram showing transverse magnetisation (T_2).....50
Figure 2-1. Vector map of pET-23b.. 57
Figure 2-2. pET23b derivative cloning and expression.....58
Figure 2-3. S-Carboxymethylation of cysteine residues scheme.....66
Figure 2-4. Typical secondary structure signals obtained by circular dichroism.. 71
Figure 2-5. Chemical shift patterns for $^{13}C\alpha$ and $^{13}C\beta$ resonances for all amino acids. 75
Figure 3-1. 12% SDS-PAGE of the b domain purification using IMAC. 81
Figure 3-2 Chromatogram of b domain using a Source 30Q column. 81
Figure 3-3. 12% SDS-PAGE of anion-exchange fractions of b domain.....82
Figure 3-4. Positive ion mass spectrum of the b domain. 83
Figure 3-5. A Far UV CD spectrum for the b domain.....84
Figure 3-6. Flow diagram of 1 mL sample treatment for SDS-PAGE analysis.....85
Figure 3-7. 12% SDS-PAGE of 1 mL b domain test induction samples. 86
Figure 3-8. 12 % SDS-PAGE of 1 mL b domain test induction samples. 86
Figure 3-9. 16% SDS-PAGE of ^{15}N b domain applied to IMAC and Anion-exchange fractions 88
Figure 3-10. 16% SDS-PAGE of anion-exchange fractions of ^{15}N-labelled b domain.....89
Figure 3-11. Chromatogram of ^{15}N b domain using a Source 30Q column. 89
Figure 3-12. Positive ion mass spectrum of the ^{15}N b domain.....90
Figure 3-13. 16% SDS-PAGE of bb'x applied to IMAC.. 92
Figure 3-14. Anion-exchange chromatogram of bb'x using a Source 30Q column. 93
Figure 3-15. 16% SDS-PAGE of fractions taken across the peak in chromatogram peak shown in Figure 3-14. 93
Figure 3-16. Chromatogram of bb'x using a Source 30Q 94
Figure 3-17. 16 % Native-PAGE of bb'x from fractions of the peak in 95
Figure 3-18. Superdex 75 gel filtration chromatogram of Pool 1. 95

Figure 3-19. 16 % native-PAGE of fractions taken across peak in chromatogram in Figure 3-18. 96
Figure 3-20. Superdex 75 gel filtration chromatogram of Pool 2. 97
Figure 3-21. 16% native-PAGE of fractions taken across peak in Figure 3-20 chromatogram. 98
Figure 3-22. Diagram showing the fractionation and pool designation. 99
Figure 3-23. Positive ion mass spectrum of the bb'x monomer. 100
Figure 3-24. Positive ion mass spectrum of the bb'x dimer. 101
Figure 3-25. A Far UV CD spectrum for the two pools of bb'x domain. 102
Figure 3-26. 16 % SDS-PAGE of bb'x test expression.. 103
Figure 3-27. 16% SDS-PAGE of bb'x test expression. 104
Figure 3-28. 16% SDS-PAGE of bb'x glucose test expression.....105
Figure 3-29. 16 % SDS-PAGE of ^{15}N bb'x applied to IMAC 106
Figure 3-30. Chromatogram of bb'x using a Source 30Q. 107
Figure 3-31. 16 % native-PAGE of bb'x from fractions of the peak in Figure 3-30. 107
Figure 3-32. Superdex 75 gel filtration chromatogram of Pool 1 108
Figure 3-33. 16 % native-PAGE of bb'x from fractions of the peaks in Figure 3-32.. 109
Figure 3-34. Superdex 75 gel filtration chromatogram of Pool 2.. 109
Figure 3-35. 16 % native-PAGE of bb'x from fractions of the peaks in Figure 3-34.. 110
Figure 3-36. Positive ion mass spectrum of ^{15}N bb'x monomer pool. 111
Figure 3-37. Positive ion mass spectrum of ^{15}N bb'x dimer pool. 112
Figure 3-38. 16% SDS-PAGE of the IMAC purification of $^{15}N/^{13}C$ bb'x. ..113
Figure 3-39. Positive ion mass spectrum of $^{13}C/^{15}N$ bb'x monomer pool...114
Figure 3-40. Positive ion mass spectrum of $^{13}C/^{15}N$ bb'x dimer pool. 115
Figure 3-41. 16% SDS-PAGE of the IMAC purification of ^{15}N b'x.....117
Figure 3-42. Chromatogram of ^{15}N b'x using a Source 30Q. 117
Figure 3-43. 16% native-PAGE of fractions across peaks in Figure 3-42.. 118
Figure 3-44. Superdex 75 gel filtration chromatogram ^{15}N b'x.....119
Figure 3-45. 16% native-PAGE of fractions across each peak in Figure 3-44. 119

Figure 3-46. Positive ion mass spectrum of ^{15}N b'x monomer.....120
Figure 3-47. Positive ion mass spectrum of ^{15}N b'x dimer.....121
Figure 3-48. Positive ion mass spectrum of $^{13}C/^{15}N$ b'x monomer. 123
Figure 3-49. Positive ion mass spectrum of $^{13}C/^{15}N$ b'x dimer.....124
Figure 3-50. bb' domain IMAC purification 125
Figure 3-51. Chromatogram of Source 30Q purification of bb'. 126
Figure 3-52. 16 % SDS-PAGE of anion-exchange fractions of bb' purification in Figure 3-51. 126
Figure 3-53. 16% native-PAGE analysis of fraction taken across the anion exchange chromatogram in Figure 3-51. 127
Figure 3-54. Positive ion mass spectrum of bb' monomer sample. 128
Figure 3-55. A Far UV CD spectrum for the bb' domain. 128
Figure 3-56. b' domain IMAC purification.. 129
Figure 3-57. Chromatogram of Source 30Q IEX purification of b'. 130
Figure 3-58. 12% SDS-PAGE of the fractions across the two peaks of the anion-exchange purification step. 131
Figure 3-59. Positive ion mass spectrum of b' Pool 1.....132
Figure 3-60. Positive ion mass spectrum of b' Pool 2.....132
Figure 3-61. A Far UV CD spectrum for the b' domain monomer pool.. 133
Figure 3-62. 16% SDS-PAGE of the IMAC purification of Full length PDI.....135
Figure 3-63. Full length PDI Source 30Q purification. 135
Figure 3-64. 16% SDS-PAGE of fractions taken across peak of full length PDI Source 30Q purification.....136
Figure 3-65. 16% native-PAGE of Full length PDI.....136
Figure 3-66. A Far UV CD spectrum for the Full length PDI domain. 137
Figure 3-67. 16% native-PAGE of various domain combinations. 140
Figure 3-68. Analytical ultra-centrifugation data analysed using SEDFIT.....142
Figure 3-69. 16% SDS-PAGE of bb'x Pool 2 'dimer' of varying concentrations under reduced and no-reduced conditions. 144
Figure 3-70. 16% SDS-PAGE of carboxymethylated bb'x under reduced and non-reducing conditions.....145

Figure 3-71. 16% native-PAGE of bb'x with different percentage of β-Mercaptoethanol in the loading buffer.....146
Figure 4-1. Schematic diagram of adaptation of bb'x *E.coli* cells to 96% D_2O.....151
Figure 4-2. 16 % SDS-PAGE of bb'x 96% D_2O test expression.. 152
Figure 4-3. 16 % SDS-PAGE of bb'x 86% D_2O test expression.. 153
Figure 4-4. 16% SDS-PAGE of bb'x 96% D_2O shock test expression.....155
Figure 4-5. 16% SDS-PAGE of bb'x 86% D_2O shock test expression.....156
Figure 4-6. 16% SDS-PAGE of 86% D_2O minimal medium expression and IMAC purification. 157
Figure 4-7. Positive ion mass spectrum of 86% D_20 shock induced bb'x ..158
Figure 4-8. 16% SDS-PAGE of 96% D_2O minimal medium expression IMAC purification. 159
Figure 4-9. Positive ion mass spectrum of 96% D_20 shock induced bb'x ..160
Figure 4-10. 16% SDS-PAGE of the affinity column protein from un-adapted cell shock induction in 96% D_2O.. 161
Figure 4-11. Positive ion mass spectrum of the affinity column eluted protein from un-adapted cell shock induction in 96% D_2O.....162
Figure 4-12. Growth curves of cultures grown in 96% D_2O minimal medium.....164
Figure 4-13. 5% Celtone ^2H 40 mL test expression.....164
Figure 4-14. Positive ion mass spectrum of the affinity column eluted protein from 5% celtone ^2H supplemented medium.....165
Figure 4-15. 10% Celtone ^2H 40 mL test expression.....166
Figure 4-16. 10% Spectra9 40 mL simple induction test expression.. 168
Figure 4-17. Positive ion mass spectrum of 10% Spectra9. 168
Figure 4-18. 10% Spectra9 40 mL resuspended shock induction test expression. 169
Figure 4-19. Positive ion mass spectrum of 10% Spectra9 40 mL shock induction test expression.....170
Figure 4-20. 16% SDS-PAGE 96% D_2O supplemented with 10% Spectra9. 172
Figure 4-21. Chromatogram of triple labelled bb'x using a Source 30Q column with a 0-100 % gradient of 1M NaCl.....173

Figure 4-22. 16% Native-PAGE of samples taken across the peak in Figure 4-21. 173
Figure 4-23. Anion-exchange chromatograph of Pool 2 174
Figure 4-24. Anion-exchange 174
Figure 4-25. 16% native gel of Pool 1 174
Figure 4-26. 16% native gel of Pool 2. 174
Figure 4-27. Positive ion mass spectrum of triple-labelled bb'x Pool 1....175
Figure 4-28. Positive ion mass spectrum of triple-labelled bb'x Pool 2....176
Figure 5-1. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-b' domain. 180
Figure 5-2. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb' domain. 181
Figure 5-3. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb'x domain Pool 1 182
Figure 5-4. $^{15}N/^{1}H$ HSQC spectrum of unfractionated ^{15}N-b'x domain....184
Figure 5-5. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb'x domain fractionated monomer sample. 185
Figure 5-6. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb'x domain fractionated dimer sample. 186
Figure 5-7. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-b'x domain monomer sample. 188
Figure 5-8. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-b domain. 189
Figure 6-1. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 25°C. 193
Figure 6-2. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 35°C after 4 days. 195
Figure 6-3. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 40°C after 30 minutes. 196
Figure 6-4. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 40°C after 22 hours. 197
Figure 6-5. $^{15}N/^{1}H$ HSQC spectrum of a new $^{13}C/^{15}N$-bb'x domain sample at 40°C 198
Figure 6-6. Typical ^{15}N plane (115.35 ppm) of CBCA(CO)NH (left) and HNCACB (right) spectra. 199
Figure 6-7. CBCA(CO)NH and HNCACB spectra of bb'x at 40°C. 200
Figure 6-8. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 40°C. 202
Figure 6-9. $^{15}N/^{1}H$ HSQC spectrum of $^{2}H/^{13}C/^{15}N$-bb'x domain at 40°C. ...204
Figure 6-10. HNCA and HN(CO)CA spectra of bb'x at 40°C 205

Figure 6-11. Fully assigned bb'x HSQC at 40°C.. 207
Figure 6-12. Fully assigned b'x HSQC at 25°C.. 207
Figure 6-13. Schematic diagram of the assignment method using the different experimental NMR data collected. 209
Figure 7-1. Zoomed in image of bb'x HSQC at 40°C.....211
Figure 7-2. Diagram showing the redefinition of species observed through various stages of characterisation and analysis. 212
Figure 7-3. $^{15}N/^{1}H$ HSQC overlaid spectrum of a $^{13}C/^{15}N$-bb'x domain sample at 40°C 214
Figure 7-4. Chemical shift difference of residues assigned in the 35°C from the 40°C spectrum.....214
Figure 7-5. Chemical shift difference of residues assigned in the 25°C from the 35°C spectrum.....215
Figure 7-6. Chemical shift perturbation identified in the b domain by comparison of the b domain HSQC and the bb'x HSQC. 217
Figure 7-7. Two viewpoints of the space-fill b domain structure.. 217
Figure 7-8. Chemical shift perturbation identified in the b'x domain by comparison of the b'x domain HSQC and the bb'x HSQC.....218
Figure 7-9. Two viewpoints of the space-fill b'x domain structural model with perturbations.....219
Figure 7-10. Diagram showing the numbered bb'x sequence, with the TALOS predictions and actual b domain secondary structure. 221
Figure 7-11. HSQC spectrum of the b domain H/D exchange after 5 minutes. 222
Figure 7-12. HSQC spectrum of the b domain H/D exchange after 21 hours.....223
Figure 7-13. The b domain ribbon diagram with the slow-exchanging 22 residues highlighted in red.....224
Figure 7-14. HSQC spectrum of the bb'x domain H/D exchange after 5 minutes.. 225
Figure 7-15. HSQC spectrum of the bb'x domain H/D exchange after 21 hours.....226
Figure 7-16. 2D-HSQC collected at 40°C of ^{15}N-bb'x with ^{1}H saturation on and off.....227

Figure 7-17. Graph showing NOEs calculated for bb'x residues. 228
Figure 7-18. Typical T_1 decay for 5 residues using peak volume against T_1 time delays. 230
Figure 7-19. R_1 rates for bb'x residues. 230
Figure 7-20. R_2 rates for bb'x residues. 231
Figure 7-21. Plot of experimental T_1 against T_2 rates. 232

LIST OF TABLES

Table 1. Summary of protein information for 15 human ER resident-PDI-family members and three PDI-like proteins from other species.. 30

Table 2. Minimal medium nutrient amounts for a 1 litre of culture. 63

Table 3. Composition of 12 and 16 % separation gel and stacking gel.. 64

Table 4. Table summarising observations during purification and storage of various domain combinations. 139

Table 5. Summary table of protein yields from 800mL expression of various domains and domain combinations. 148

Table 6. Summary of peak width analysis of different domain combinations. 179

Table 7. $^{13}C/^{15}N$-bb'x domain HNCACB quality assessment. 194

ABBREVIATIONS

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
Å	angstrom
A	adenine
A_{595}	absorbance at wavelength 595 nm
A_{600}	absorbance at wavelength 600 nm
APS	ammonium persulfate
AU	absorbance unit
B-ME	β-mercaptoethanol
bp	base pair
BiP	binding protein
BPTI	bovine pancreatic trypsin inhibitor
BSA	bovine serum albumin
Buffer A	20 mM Sodium Phosphate pH 7.3
Buffer B	20 mM Sodium Phosphate, 0.5 M NaCl pH 7.3
C	cytosine
Cα	alpha carbon of an amino acid
Cβ	beta carbon of an amino acid
CD	circular dichroism
cm	centimetre
C-terminus	carboxy terminus
CV	column volume
D_2O	deuterium oxide
Da	dalton
DNA	deoxyribonucleic acid
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
ES	electrospray mass spectrometry
et al.	*Et alia* (and others)

ER	endoplasmic reticulum
ERO1	endoplasmic reticulum oxidoreductin 1
G	guanine
GAPDH	D-glyceraldehyde-3phosphate dehydrogenase
GSH	reduced glutathione
GSSG	oxidised glutathione
HCl	hydrochloric acid
H/D	hydrogen/deuterium exchange
HetNOE	heteronuclear nuclear overhauser effect
His-tag	hexa-histidine sequence
HN	amide proton
HPLC	high performance liquid chromatography
IMAC	immobilised metal affinity chromatography
IPTG	isopropyl-β-D-thiogalactopyranoside
kb	kilobase
KCl	potassium chloride
kDa	kilodalton
K_m	michaelis constant
L	litre
LB	luria-bertani
m/z	mass to charge ratio
M	molar (grams per litre)
mAU	milli absorbance unit
mg	milligram
$MgCl_2$	magnesium chloride
min	minutes
mL	millilitre
mM	millimolar
M_r	molecular weight
mRNA	messenger ribonucleic acid
$mS.cm^{-1}$	millisiemens per centimetre
MWCO	molecular weight cut-off
NAD^+	nicotinamide adenine dinucleotide (oxidised)
NADH	nicotinamide adenine dinucleotide (reduced)

$NADP^+$	nicotinamide adenine dinucleotide phosphate (oxidised)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
ng	nanogram
Ni	nickel
NMR	nuclear magnetic resonance
NOE	nuclear overhauser effect
N-terminus	amino terminus
OD	optical density
°C	degrees Celsius
P4H	prolyl 4-hydroxylase
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PDI	protein disulphide isomerase
PDIp	pancreas specific protein disulphide isomerase
pH	$\log_{10} [H^+]$
pmol	picomole
ppm	parts per million
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
sec	second
T	thymine
Tris	tris (hydroxymethyl) aminomethane
TAE	tris acetate ethylenediaminetetraacetic acid
TEMED	tetramethyl-ethylenediamine
TBS	tris buffered saline
TOCSY	total correlation spectroscopy
tRNA	transfer ribonucleic acid
µg	microgram
µL	microlitre
µM	micromolar

µmol	micromole
UPR	unfolded protein response
UV	ultraviolet
V_{max}	maximal velocity
V	volts
v/v	volume to volume ratio
w/v	weight to volume ratio
x g	centrifugal force

Other abbreviations are explained where appropriate in text.

Single and three letter symbols for amino acids

Amino Acid	Three Letter Code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Chapter 1. General Introduction

1.1. Introduction

This Chapter presents an overview of this study. It will summarise protein folding in the endoplasmic reticulum (ER) and how protein disulphide isomerase (PDI), a multifunctional enzyme involved in thiol-disulphide exchange acts as a foldase and chaperone; facilitating protein folding. It will describe structural and functional information available for PDI and related proteins. Due to the intransigent nature of the **b'** domain of PDI, structural determination by x-ray crystallography and nuclear magnetic resonance (NMR) has previously been unsuccessful. In addition, this Chapter will describe the aims of this project to overcome the difficulties in obtaining structural data on the **b'** domain by working on the **bb'x** and **b'x** domain constructs. Preliminary data, discussed later in detail, suggests that the **b'** domain is more stable in these constructs.

Protein synthesis and the flow of cellular information follow a pathway with several critical stages starting with transcription in the nucleus of specific DNA segments by RNA polymerase to produce messenger RNA (mRNA). These mRNA molecules are then translated in the cytoplasm, with the aid of transfer RNA (tRNA) into proteins. These newly synthesised protein molecules are not immediately physiologically functional; they require folding into their native state. Proteins destined for secretion to the outer membrane and compartments in the secretory pathway such as the Golgi, must first enter the endoplasmic reticulum (ER) in an unfolded state and only exit once they have correctly folded and assembled. It is estimated that 33% of all human proteins fold in the ER (Chen *et al.*, 2005). The ER is highly regulated and has several defined functions such as steroid production (Miller, 2005), but the main function is protein folding and post-translational modification which takes place in a specialized compartment, the lumen of the ER. The ER therefore contains a wide variety of catalysts for

post-translational modifications such as N-glycosylation (Helenius and Aebi, 2004) and protein folding such as disulphide bond formation (Sevier and Kaiser, 2002).

The formation of disulphide bonds requires an oxidising environment (Raines, 1997), the ER provides this environment where the reduction potential ($E^{o'}$) is -0.18V, hence allowing the formation of native disulphide bonds. Disulphide bonds are covalent linkages between two cysteine residues in proteins and their main function is protein structure stability in the extracellular environment. Since any cysteine residue can, in principle, form a disulphide bond with any other inter or intra-molecular cysteine during the process of protein folding, the rate-limiting step is correct disulphide bond formation (Creighton *et al.*, 1995). Protein folding in the context of the formation of native disulphide bonds is carried out by the ubiquitously expressed protein disulphide isomerase (PDI) which can increase the rate of native disulphide bond formation by a factor of between 1000 to 6000 by catalysis of disulphide bond formation and interchange (Freedman *et al.*, 1994).

In mammalian cells, proteins begin to fold co-translationally in the ER and folding is completed post-translationally. Proteins are synthesised by ER-associated ribosomes and are translocated through the ER membrane via a translocon believed to be the Sec61 protein complex as shown in Figure 1-1, part a. This combination of processes translation-translocation-sequential folding immediately as proteins are synthesised, has been shown to enhance the folding efficiency of proteins, particularly with multi-domain proteins (Netzer and Hartl, 1997).

Newly synthesised proteins transported into the ER are retained in the ER due to their interactions with chaperones (Ellgaard and Helenius, 2001). Unfolded or partially folded proteins in ER have exposed hydrophobic regions and so are likely to be more prone to aggregation. Small and growing aggregates are

potentially cytotoxic and so a quality control method of preventing their build up is essential (Ellgaard and Helenius, 2003b). Chaperones are believed to be able to recognise proteins in this state, interact with them and prevent their aggregation (Fewell *et al.*, 2001). A two-chaperone system exists to prevent their build up and facilitate the folding of mis-assembled, aggregation-prone and newly-synthesised unfolded proteins.

The first system involves lectin-like chaperones, calnexin (CNX) and calreticulin (CRT) as shown in Figure 1-1, part b. As newly synthesised glycoproteins enter the ER they are modified by N-linked glycans consisting of two N-acetylglucosamines, nine mannose and three glucose molecules. Two of the glucose molecules are trimmed by glucosidase I and II (Kornfeld and Kornfeld, 1985), the single glucose glycoprotein is then a substrate for CNX and CRT which bind. CNX and CRT facilitate folding by preventing aggregation and retaining the nascent protein in the ER, while the actual functional folding is carried out with the aid of accessory proteins such as UDP-glucose:glycoprotein glucosyltransferase (UGGT) and ERp57, a member of the PDI family (Parodi, 1991). The removal of the final glucose on the nascent protein leads to the removal of the CNX and CRT binding site and provided the nascent protein is fully folded, the nascent protein is then transported to the Golgi (Hebert *et al.*, 1995) as shown in Figure 1-1, part c.

The second system involves ER chaperone BiP which recognises hydrophobic regions of unfolded proteins (Flynn *et al.*, 1991). BiP belongs to the ER Hsp70 family of proteins which all bind ATP and ADP, which regulate the binding and release of nascent protein molecules in a process called the ATPase cycle (Bukau and Horwich, 1998). The cycles of binding and release during the Hsp70 ATPase cycle promote the folding of the nascent protein to a more folded state, with the folding occurring during the release cycle and the binding preventing aggregation or inappropriate interactions (Hendershot *et al.*, 1996).

When proteins are released by CNX, CRT or BiP and if the secretory protein is correctly folded, then the secretory protein is transported out of the ER via transport vesicles and continues on the secretory pathway to the Golgi as shown in Figure 1-1, part c. If the protein is a glycoprotein then exit from the ER is facilitated by ER-Golgi intermediate compartment (ERGIC-53) (Schrag *et al.*, 2003). Alternatively if the proteins released by the chaperones are mis-assembled then they are removed from the folding pathway and targeted to the ER-associated degradation (ERAD) pathway. The upstream signals targeting mis-assembled proteins to downstream degradation remain unclear. BiP is known to associate with the protein translocation channel (Sec61-containing translocon structure) during import into the ER. (Wiertz *et al.*, 1996). BiP forms a seal on the luminal side of the ER channel during import (Hamman *et al.*, 1998). This same translocation channel involved in protein import has been identified as the point of exit for mis-assembled proteins targeted for ER export. PDI has recently been shown to interact with BiP whilst bound to the translocation channel. Therefore a proposed method of protein export involves PDI delivering the mis-assembled protein to the translocon, where a PDI-BiP-mis-assembled protein complex forms on the luminal side of the ER which triggers an opening in the channel, then the mis-assembled protein can pass through and out into the cytosol (Gillece *et al.*, 1999), as shown in Figure 1-1, part d. During the process of exporting from the ER, mis-assembled proteins are polyubquitinated on the cytosolic surface of the ER membrane (Hershko and Ciechanover, 1998) and then degraded by the cytosolic 26S proteasome (Oberdorf *et al.*, 2001).

During the process of ERAD, high levels of mis-assembled proteins in the ER also trigger the unfolded protein response (UPR), which is a coordinated response to reduce the levels of mis-assembled proteins as shown in Figure 1-1, part e. Three main ER-transmembrane sensors IRE1, PERK and AFT6 act in a synergic fashion (Kaufman, 2002), whereby AFT6 up-regulates the transcription of ER proteins and chaperones, Ire1 indirectly induces transcription factors that

facilitate ERAD, and PERK transiently attenuates translation, limiting protein load. This coordinated response is an efficient method of maintaining the balance of retaining and degrading harmful proteins and yet continuing the export of biologically active and essential proteins (Sitia and Braakman, 2003). If this response fails to clear the ER of the high concentration of harmful mis-assembled proteins, then apoptosis is induced (Otomo *et al.*, 1999).

A similar response occurs in the cytosol, whereby the accumulation of mis-assembled proteins in the cytosol triggers the heat-shock response which involves the transcriptional up-regulation of Hsp70 and other cytosol chaperones (Morimoto, 1998).

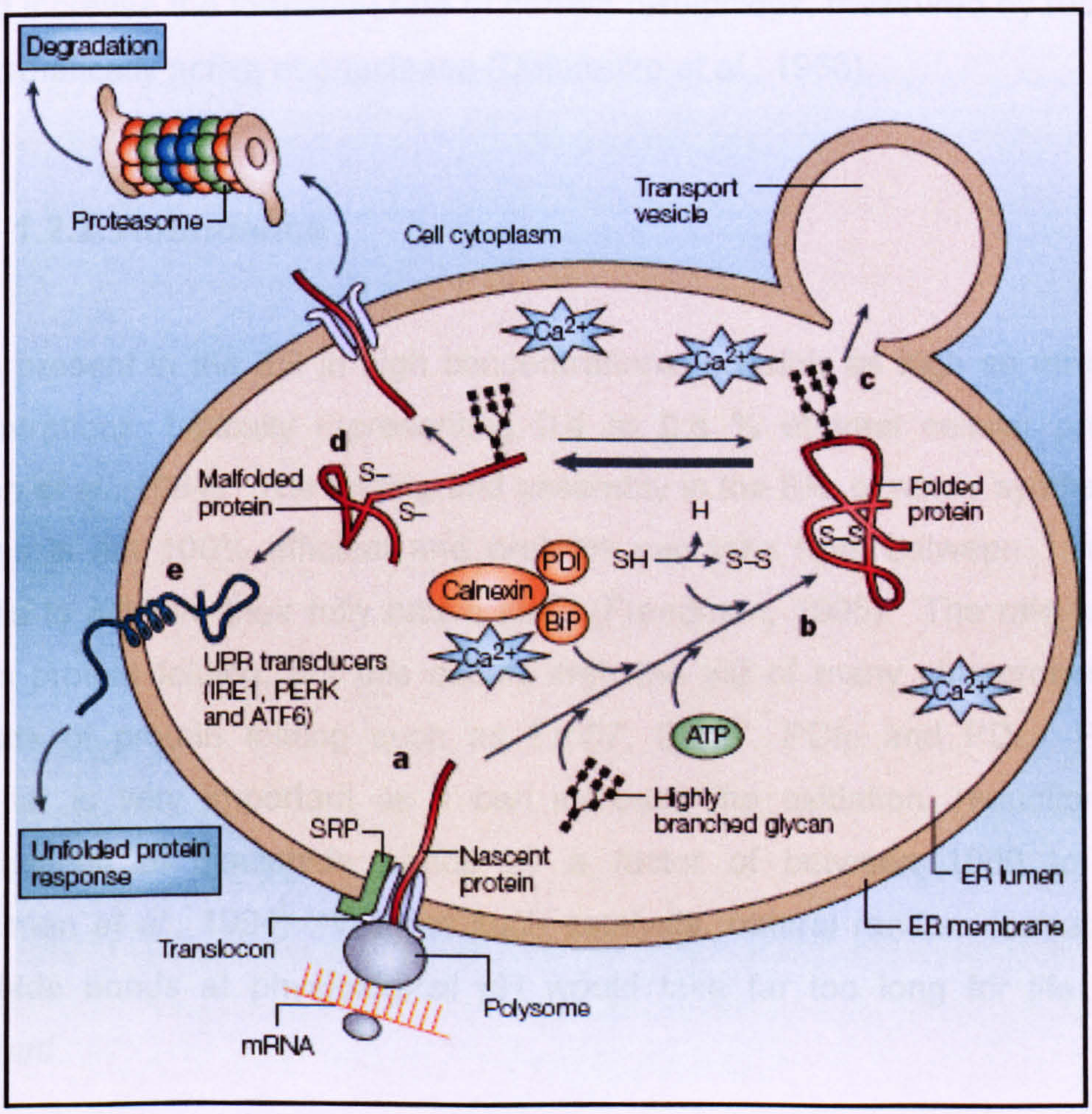

Figure 1-1. Schematic diagram of protein folding and degradation in the endoplasmic reticulum (taken from Ma and Hendershot, 2004).

1.2. Protein Disulphide Isomerase

1.2.1. Activity

PDI (E.C. 5.3.4.1) was the first identified catalyst of protein folding (Delorenzo *et al.*, 1966; Freedman, 1984), recognised by Anfinsen in 1964 to be a catalyst of disulphide bond formation. By studying the components involved in enhancing the rate of reactivation of reduced ribonuclease in rat liver lysates, this group found that PDI did not contribute to the net oxidation of reduced ribonuclease but did increase the formation rate of correct disulphides; measured by recovery of enzymatically active ribonuclease (Delorenzo *et al.*, 1966).

1.2.2. Abundance

PDI is present in the ER in high concentrations, possibly as high as millimolar concentrations, typically representing 0.4 to 0.8 % of total cellular proteins (Hillson *et al.*, 1984). The folding and assembly in the ER, of newly synthesised proteins is not 100% efficient and proteins can take from between 1 to 100 minutes to achieve their fully native state (Freedman, 1995). The rate-limiting step is protein folding, but this occurs with the aid of many chaperones and catalysts of protein folding such as Erp57, Erp27, PDIp and PDI. PDI in particular is very important as it can increase the oxidation, reduction and isomerisation of disulphide bonds by a factor of between 1000 to 6000 (Freedman *et al.*, 1994). Without such catalysts, natural random formation of disulphide bonds at physiological pH would take far too long for life to be sustained.

1.2.3. PDI location

PDI is found in a variety of species, tissues and organs; species include bovine (Delorenzo *et al.*, 1966), chicken, pigeon and pig (Venetianer and Straub, 1963). PDI is also found in a range of tissue such as the liver and pituitary gland, which have hosted the highest concentrations of PDI in mammalian species. PDI cellular localisation also appears to be broad and not just localised to the ER. Interestingly, PDI has been linked to extracellular locations, such as the cell-adhesion protein retina cognin (Pariser *et al.*, 1998), which is reported to be involved in cell recognition and neuronal differentiation of embryonic retina cells. This protein has been speculated to be a truncated version of PDI from which the KDEL ER retention signal is removed (Phillips *et al.*, 1997; Pariser *et al.*, 1998). PDI is also found as a subunit of prolyl-4-hydroxylase (P4-H) and microsomal triglyceride transferase (MTP), discussed further later in sections 1.2.9 and 1.2.10.

1.2.4. Redox-Isomerase function

Research on the refolding of denatured ribonuclease A, showed that disulphide bond formation could occur spontaneously (Anfinsen, 1973). It was observed that disulphide bond formation could occur much slower in air than in the cell, suggesting the involvement of enzymatic catalysis for oxidative protein folding. This led to the identification of the enzymatic catalyst of oxidative refolding (disulphide bond formation), PDI (Goldberger *et al.*, 1963). After exhaustive investigations the *in vitro* function of PDI has been shown to be the catalysis of the oxidation, reduction and isomerisation of disulphide bonds. Another possible method of direct oxidation by molecular oxygen is possible but without any catalysts the reaction would be too slow.

1.2.4.1. Ero1

Ero1 (ER oxidoreductin) is a 65 kDa flavoenzyme, associated with the luminal face of the ER membrane. Several Ero1 proteins from a number of species have been characterised including Human (Cabibbo *et al.*, 2000) and *Saccharomyces cerevisiae* (Frand and Kaiser, 1998). Ero1p is a yeast membrane associated ER protein homologue of human Ero1. Ero1p was shown to be involved in producing oxidising equivalents necessary for protein disulphide bond formation in the ER, since intra-molecular disulphides were not present in secretory proteins of Ero1p knockout cells (Frand and Kaiser, 1998). Interestingly, mixed disulphide linked complexes between Ero1p and PDIp (yeast PDI homologue) were found, suggesting the PDIp is an important intermediate in the transfer of disulphide bonds between Ero1p and secretory proteins (Frand and Kaiser, 1999).

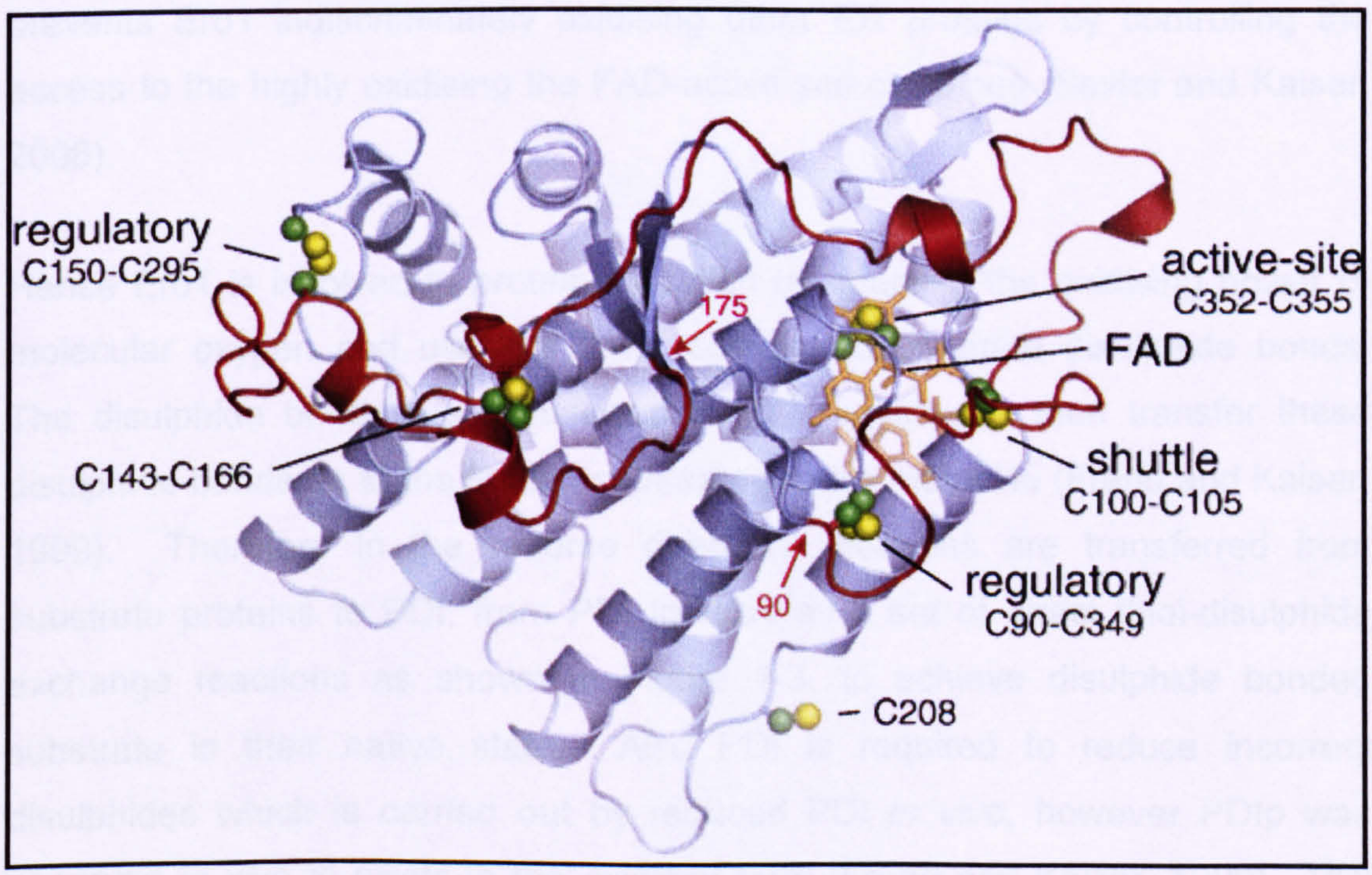

Figure 1-2. Yeast Ero1 structure, taken from Sevier and Kaiser, 2000b. Green and yellow spheres highlight the cysteine side chains, orange sticks highlight the FAD cofactor. The polypeptide chain which contains the shuttle cysteines (residues 90-175) are highlighted red.

Ero1 has two pairs of catalytic cysteines, the shuttle cysteines (C100-105) and active site cysteines (C352-C355). There are also two pairs of regulatory cysteines (C90-C349, C150-C295) as highlighted in Figure 1-2 (Frand and Kaiser, 2000b). Ero1 activity and flexibility of the polypeptide chain around the shuttle cysteines is believed to be controlled by the regulatory cysteine pairs (Sevier *et al.*, 2007).

Ero1 catalysis works by firstly direct oxidation of the PDI by the shuttle cysteine pair and then re-oxidation of the now reduced shuttle cysteine pair by internal dithiol disulphide exchange with the active site cysteine pair (Frand and Kaiser, 2000b). Electrons are then transferred from active site cysteine pair to the flavin cofactor and molecular oxygen to produce hydrogen peroxide, resulting in the re-oxidation of the active site cysteine pair (Tu *et al.*, 2000). This process using the electron shuttling between two cysteines and direct oxidation of PDI, prevents Ero1 indiscriminately oxidising other ER proteins by controlling the access to the highly oxidising the FAD-active site cysteines (Sevier and Kaiser, 2008).

Hence Ero1 is involved in protein oxidation by coupling the oxidising power of molecular oxygen and uses its flavin cofactor to generate disulphide bonds. The disulphide bonds are transferred to PDI, which can then transfer these disulphide bonds on to the folding substrate protein molecule (Frand and Kaiser, 1999). Therefore in the reverse direction, electrons are transferred from substrate proteins to PDI, from PDI to Ero1 as a set of direct thiol-disulphide exchange reactions as shown in Figure 1-3, to achieve disulphide bonded substrate in their native state. Also PDI is required to reduce incorrect disulphides which is carried out by reduced PDI *in vivo*, however PDIp was observed *in vivo* to exists in the oxidised form (Frand and Kaiser, 1999). PDI could be reduced by GSH, so facilitate reduction and lead to isomerisation of disulphide bonds in substrate proteins, as shown in Figure 1-3. Hence there

appears to be some functional cross-talk between PDI, Ero1 and glutathione pathways (Frand *et al.*, 2000a). This model only considers direct oxidoreductase and does not address the formation of incorrect disulphides and their subsequent isomerisation to achieve the correct final fully folded native state, which has been shown to be the rate-determining step *in vitro* and in the cell (Land *et al.*, 2003). Importantly PDI family members are the only enzymes which catalyse these complex rate limiting isomerisation reactions (Weissman and Kim, 1993).

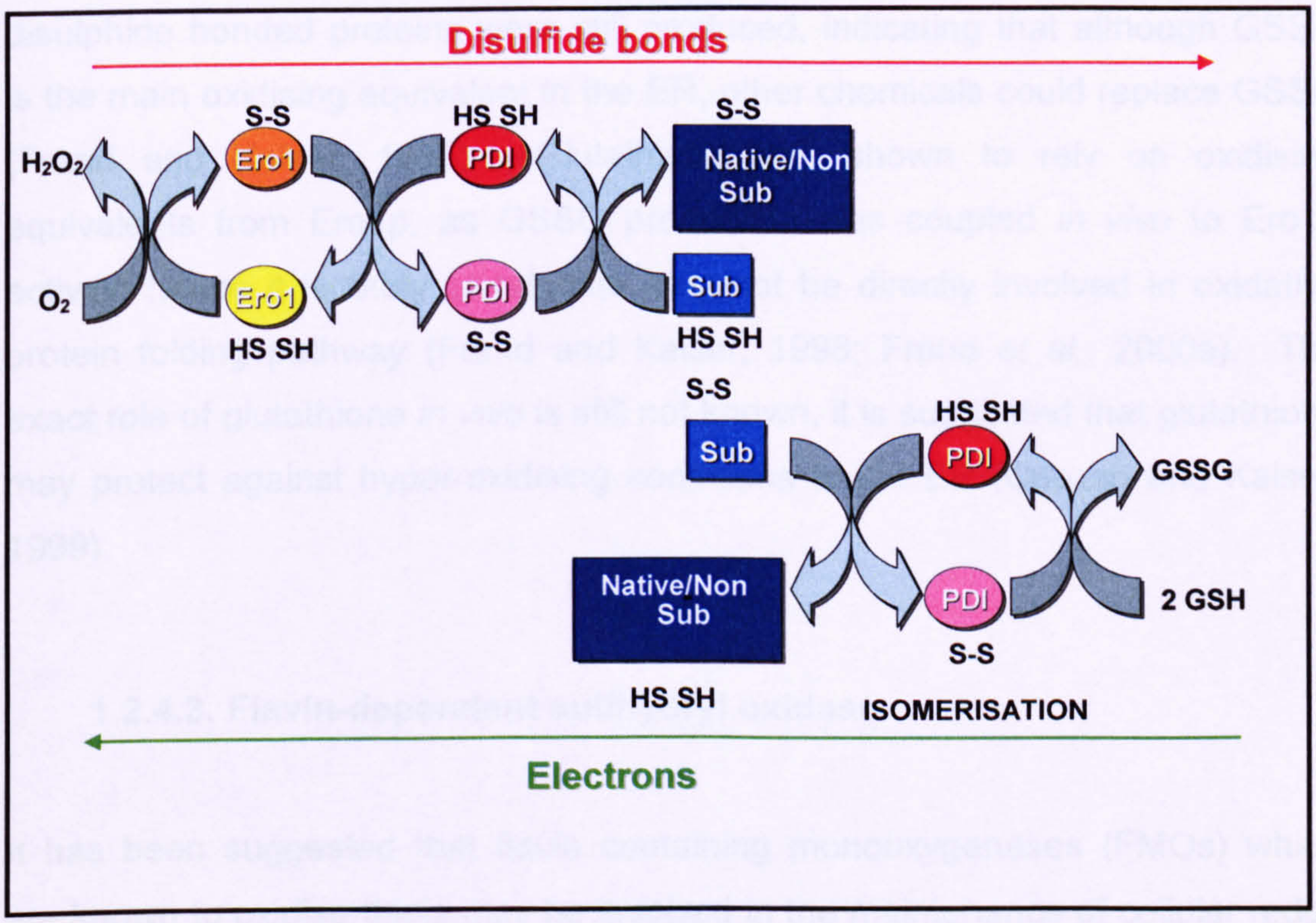

Figure 1-3. A Graphical representation of the proposed electron flow, disulphide bond formation pathway for the oxidation and isomerisation of protein substrates.

1.2.4.2. Oxidised glutathione

Native disulphide bond formation requires an oxidising environment and in eukaryotes this is provided by the ER. Glutathione is a major small molecule redox buffer in the ER. The redox potential relies on the ratio of 1/1 or 3/1 GSH

(reduced) (Bass *et al.*, 2004) and GSSG (oxidised) (Hwang *et al.*, 1992) forms in the ER with a total concentration >1mM. The cytosolic GSH/GSSG ratio is 30/1 to 100/1 and a total concentration of approximately 10mM; hence the ER is 20 to 100 times more oxidative than the cytosolic environment. It is not known how exactly the higher oxidative state of the ER is maintained especially since, it is suggested that there is a preferential transport of GSH into the ER from the cytosol (Banhegyi *et al.*, 1999).

It was shown in mutant cells unable to produce glutathione (GSH) that disulphide bonded proteins were still produced, indicating that although GSSG is the main oxidising equivalent in the ER, other chemicals could replace GSSG (Frand and Kaiser, 1998). Glutathione was shown to rely on oxidising equivalents from Ero1p, as GSSG production was coupled *in vivo* to Ero1p activity, hence it is likely glutathione may not be directly involved in oxidative protein folding pathway (Frand and Kaiser, 1998; Frand *et al.*, 2000a). The exact role of glutathione *in vivo* is still not known, it is suggested that glutathione may protect against hyper-oxidising conditions in the ER (Cuozzo and Kaiser, 1999).

1.2.4.3. Flavin-dependent sulfhydryl oxidase

It has been suggested that flavin containing monooxygenases (FMOs) which are known to oxidise thiols may be involved in the maintenance of cellular redox balance (Ziegler and Poulsen, 1977). It was recently shown with the yeast homolog of FMO, yFMO, that in yFMO inactive cells the ER were unable to maintain the optimal oxidising environment for the folding of disulphide bonded proteins (Suh *et al.*, 1999). These experiments also showed that yFMO is localised to the cytoplasmic surface of the ER membrane. It is known that GSH reductase catalyses the NADPH dependent reduction of GSSG to GSH (Hwang *et al.*, 1992) and this is essentially the main system of controlling the redox

potential in the cytosol. So it appears that both GSH reductase and yFMO control the GSH/GSSG ratio in the cytosol and then by an unknown process GSH is transported into the ER.

Members of the Quiescin-sulfhydryl oxidase (QSOX) family have a thioredoxin domain and a small FAD-binding domain homologous to the yeast ERV1p protein, which are used to oxidize sulfhydryl groups to disulphides with the reduction of oxygen to hydrogen peroxide. Isolated from human lung fibroblasts, hQSOX1a (Coppock *et al.*, 1993) and other enzymes from the QSOX family have the ability to produce disulphide bonded proteins *in vitro* (Hoober *et al.*, 1999). QSOX enzymes have been shown to possess the ability to introduce disulphide bonds directly into substrate protein molecules without the requirement of any interaction with PDI or Ero1 *in vitro*, but this ability causes the rapid and indiscriminate oxidation of substrate proteins (Hoober *et al.*, 1999). The QSOX enzyme has been shown to be localised to Golgi apparatus and so is believed to be involved in the later stages of the secretory pathway, either to maintain disulphide bonds or oxidise specific substrates (Chakravarthi *et al.*, 2007). QSOX family of proteins have also been found to be distributed in a range of species and highly active secretory tissue (Thorpe *et al.*, 2002). As a result of the wide distribution of QSOX proteins it has been hypothesized that there may be new pathways of disulphide bond formation outside the eukaryotic ER (Chakravarthi *et al.*, 2007).

1.2.5. Thioredoxin fold

PDI is a member of the thioredoxin protein superfamily. Members of this superfamily have an α/β fold with an overall domain structure consisting of β- α-β-α-β-α-β-β-α folds (Martin, 1995), whereby the central core consists of a five stranded β-sheet (four are parallel except β4) which is surrounded by 4 α-helices as shown in Figure 1-4. The x-ray crystal structure of oxidised form of

thioredoxin was determined in 1990 (Katti *et al.*, 1990) and shortly after the reduced structure was determined by NMR (Forman-Kay *et al.*, 1991).

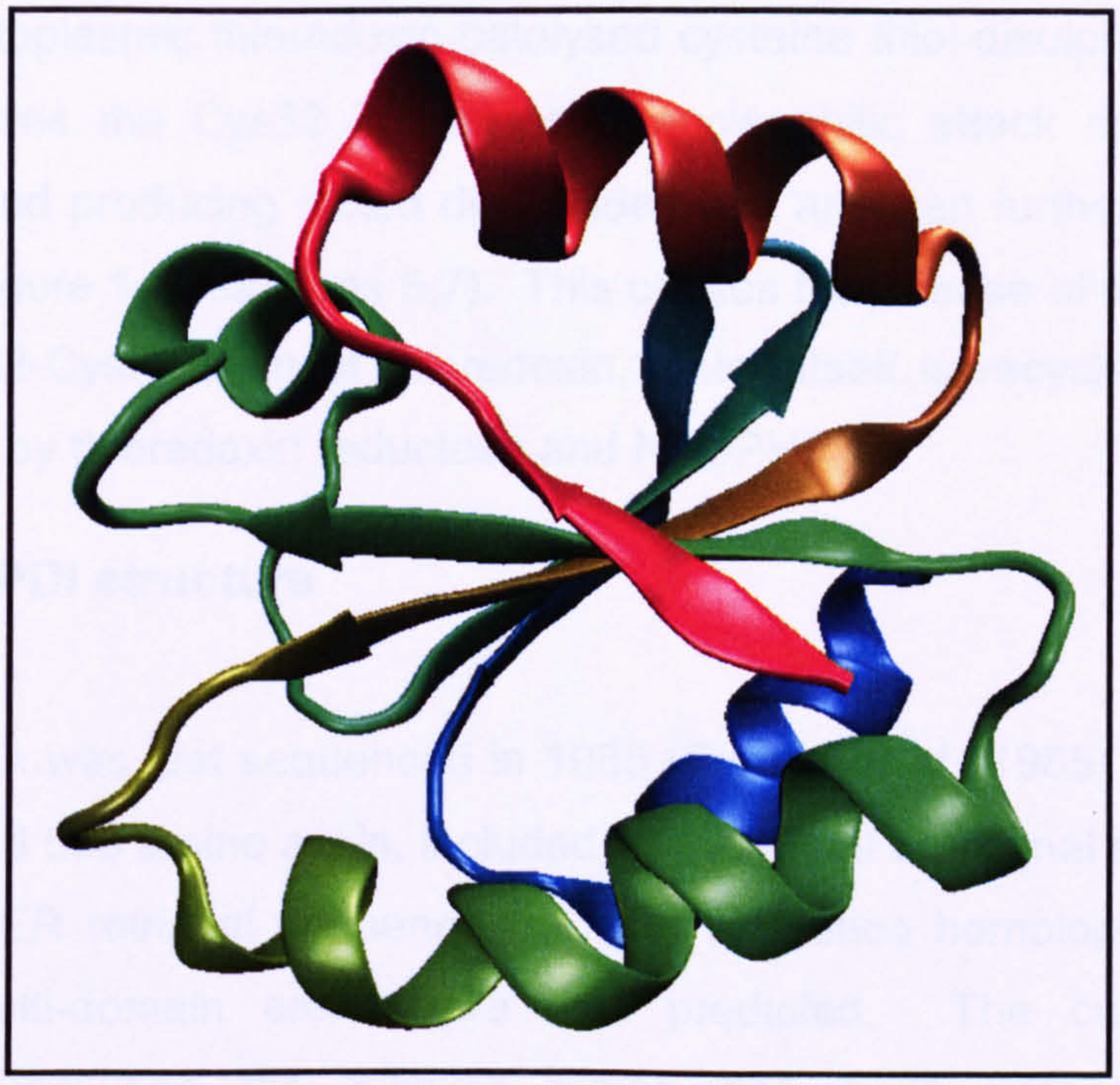

Figure 1-4. Structural diagram of Human thioredoxin (reduced form) PDB ID 1ERT.

Thioredoxin is found in all species and is a small 12 kDa protein, localised to the cytosol, nucleus and mitochondria. This enzyme is primarily a protein reductant but has been linked to stimulating cell growth, inhibition of apoptosis and is believed to play a role in a variety of human diseases including cancer, where an increased level of thioredoxin was found in many human tumour cells (Powis and Montfort, 2001). A number of protein families such as glutaredoxin, DsbA and PDI, share this archetypal thioredoxin fold and have been associated with sulfur-based redox reactions. Most of these enzymes have a conserved active site motif of Cys-X-X-Cys. Thioredoxin has these two vicinal cysteines in a Cys-X-X-Cys tetra-peptide active site motif which is exposed between β2 and α2. The C-terminal Cys35 is buried behind the N-terminal end Cys residue (Cys32), which is highly reactive since it is surface exposed, at the N-terminal end of the α2 helix, with a pKa of 7.1 (Dyson *et al.*, 1997); lower than a free cysteine which

would have a pKa of 8.7. It has been suggested that the buried partial charge on Asp26 influences Cys32 and Asp26 acts as a general acid/base for thioredoxin catalysed redox reactions (Chivers and Raines, 1997). The general method of cytoplasmic thioredoxin-catalysed cysteine thiol-disulphide exchange activity, involves the Cys32 in making a nucleophilic attack on a substrate disulphide bond producing mixed disulphides that are then further disrupted by the Cys35 (Figure 1-8 reactions 5,7). This causes the release of the disulphide-bonded Cys32-Cys35 form of thioredoxin, which itself is recycled back to its reduced form by thioredoxin reductase and NADPH.

1.2.6. PDI structure

The PDI cDNA was first sequenced in 1985 (Edman *et al.*, 1985) from rat liver, which encoded 508 amino acids, included a N-terminal ER signal sequence and a C-terminal ER retrieval sequence. Using sequence homologies within the protein a multi-domain architecture was predicted. The current domain architecture, including the **x**-linker region has been proposed using a combination of experimental data (in terms of proteolysis of native PDI and the characterisation of recombinant fragments) and bioinformatics data. The current model therefore describes 4 structural domain **a**, **b**, **b'** and **a'**, a C-terminal acidic extension and a short 19 residue **x**-linker region between domain **b'** and **a'**; as shown in Figure 1-5 (Kemmink *et al.*, 1997; Freedman *et al.*, 1998). The catalytic **a** and **a'** domains are separated by two non-catalytic **b** and **b'** domains; also the **x**-region. It is likely that PDI evolved from internal gene duplication, which over time lost sequence similarity but retained the overall domain fold. Studies on the molecular evolution of PDI using phylogenetic analysis have suggested PDI evolved from an ancestral thioredoxin-like double domain protein (Kanai *et al.*, 1998; McArthur *et al.*, 2001).

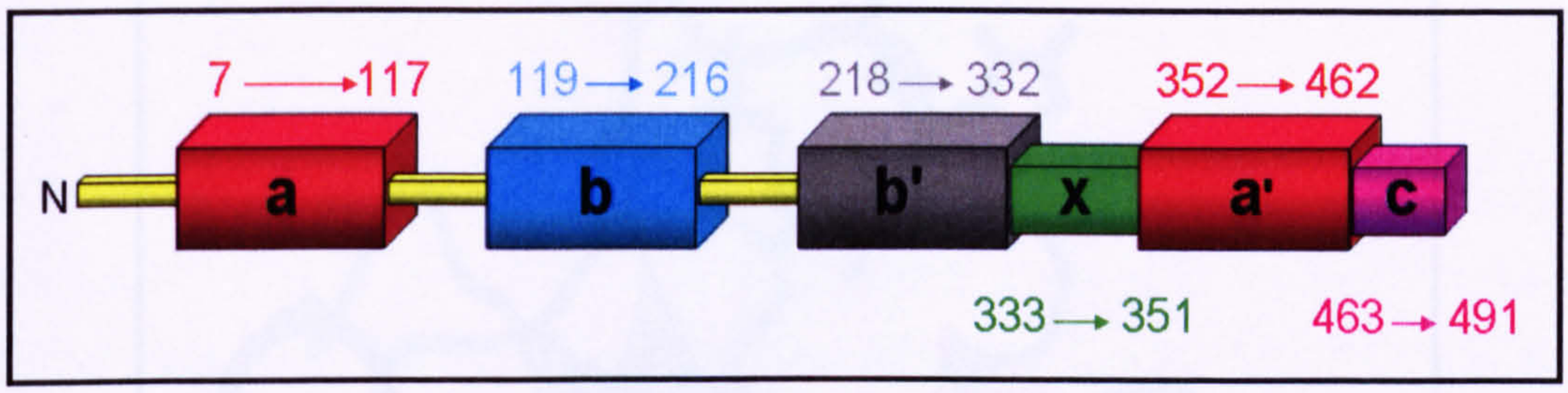

Figure 1-5. PDI domain architecture. Domain boundaries are based on NMR-derived structural information, homology modelling of b' domain and data from limited proteolysis of native PDI. Residue numbers are for mature human PDI.

1.2.6.1. a and a' domains

The **a** domain was the first domain structurally determined by NMR shown in Figure 1-6, in 1996 (Kemmink *et al.*, 1996) and as yet unpublished NMR assignments and low resolution structure of the **a'** domain has been submitted to the Protein data bank, PDB file 1X5C (Tochio *et al.*, to be published). Both **a** and **a'** domains show a significant sequence similarity to thioredoxin, 41% and 32% respectively. The homologous **a** and **a'** domains have a 49% sequence similarity and both have the Cys-X-X-Cys tetra-peptide active site motif (Cys-Gly-His-Cys in human PDI). This motif lies at the N-terminus of the α-helix2 and just prior to the β-strand 3, there is also a conserved cis-proline which is spatially juxtaposed and associated with substrate interactions in other protein family members (Kemmink *et al.*, 1996); similar features are displayed in thioredoxin.

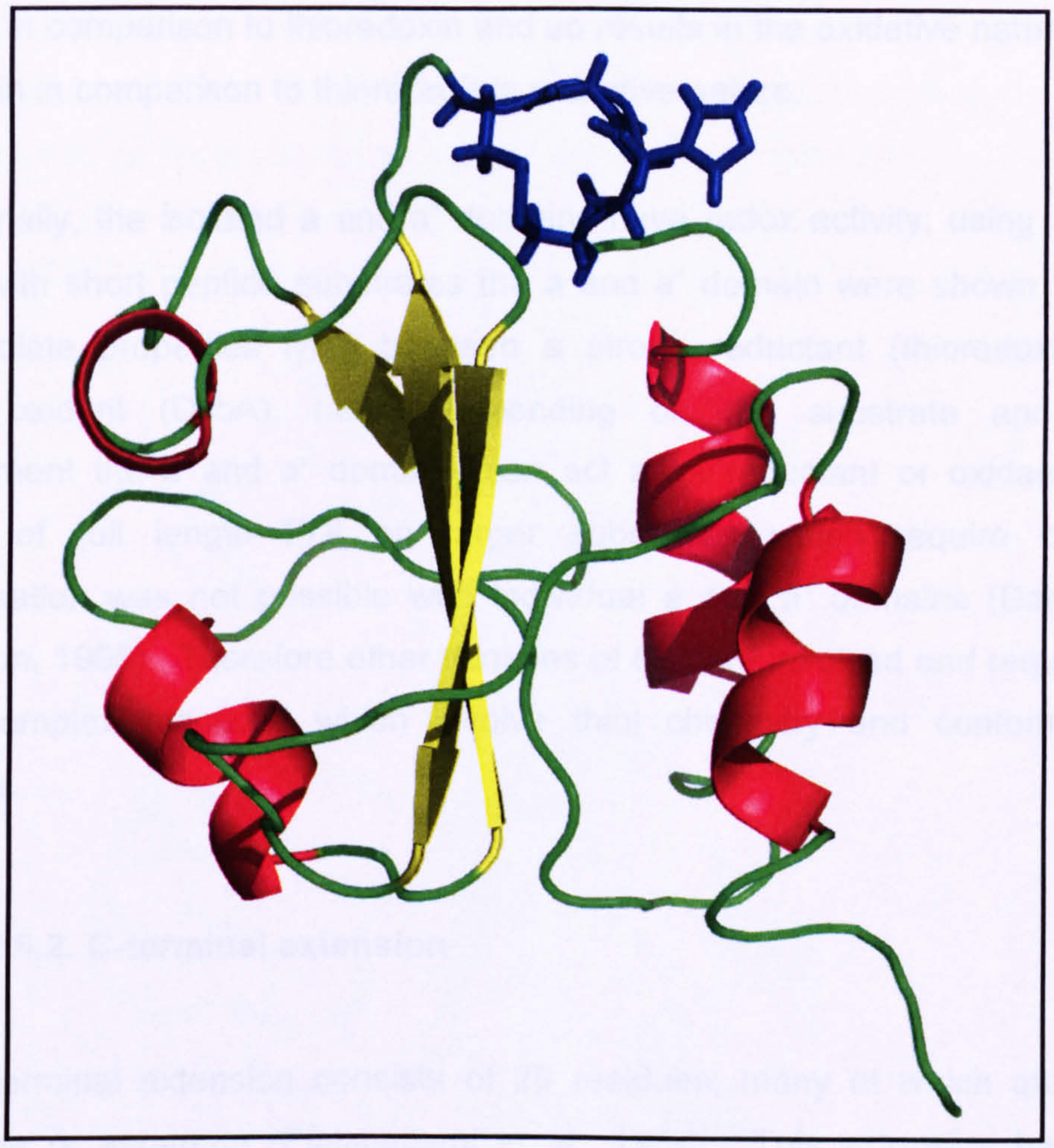

Figure 1-6. Structural diagram of the human PDI a domain (PDB code - 1MEK). α-helices shown in red, β-sheets shown in yellow and active site residues shown in blue sticks.

Furthermore the **a** domain has a buried acidic residue, Glu30, which is in an analogous position to the Asp26 residue in thioredoxin, which is believed to affect redox properties of the protein (Martin, 1995). In addition to sharing these features with the thioredoxin domain, the **a** domain also has the same overall fold. Although there is no high resolution **a'** domain structure; there do appear to be similarities with the **a** domain distinguishable using the low resolution structure. A difference between the **a** domain and thioredoxin is that the N-terminal cysteine residue in the **a** domain active site has a pKa of 4.5, which is stabilised by the active site histidine imidazole group and positive charged amino acid side chains which occur after the active site motif and before the N-terminus of the α-helix2. This results in the less stable disulphide form of the **a**

domain in comparison to thioredoxin and so results in the oxidative nature of the **a** domain in comparison to thioredoxin's reductive nature.

Functionally, the isolated **a** and **a'** domains have redox activity; using a redox assay with short peptide substrates the **a** and **a'** domain were shown to have intermediate properties lying between a strong reductant (thioredoxin) and strong oxidant (DsbA), hence depending on the substrate and redox environment the **a** and **a'** domains can act as a reductant or oxidant. The activity of full length PDI on larger substrates which require complex isomerisation was not possible with individual **a** and **a'** domains (Darby and Creighton, 1995). Therefore other domains of PDI are involved and required for more complex reactions which involve thiol chemistry and conformational changes.

1.2.6.2. C-terminal extension

The C-terminal extension consists of 29 residues, many of which are either glutamate or aspartate (Pihlajaniemi *et al.*, 1987). This extension has been reported to have a putative low-affinity, high capacity calcium ion binding site (Macer and Koch, 1988). The functional importance of this acidic extension is not known, but this extension does contain the –KDEL, ER retention sequence which is essential to retain PDI in the ER lumen.

1.2.6.3. b and b' domain

The **b** domain has little sequence similarity to thioredoxin (only 27%) but from the NMR determined structure and like the **a** and **a'** domain, the **b** domain has a thioredoxin fold as shown in Figure 1-7 (Kemmink *et al.*, 1999). But unlike the **a** and **a'** domain, the **b** and **b'** domains do not have the conserved thioredoxin–like active site residues or other redox associated residues which have been

replaced and therefore the **b** and **b'** domains do not have dithiol-disulphide oxidoreductase activity. The **b'** domain does have a high sequence similarity to the **b** domain, which is also expected to have a thioredoxin fold.

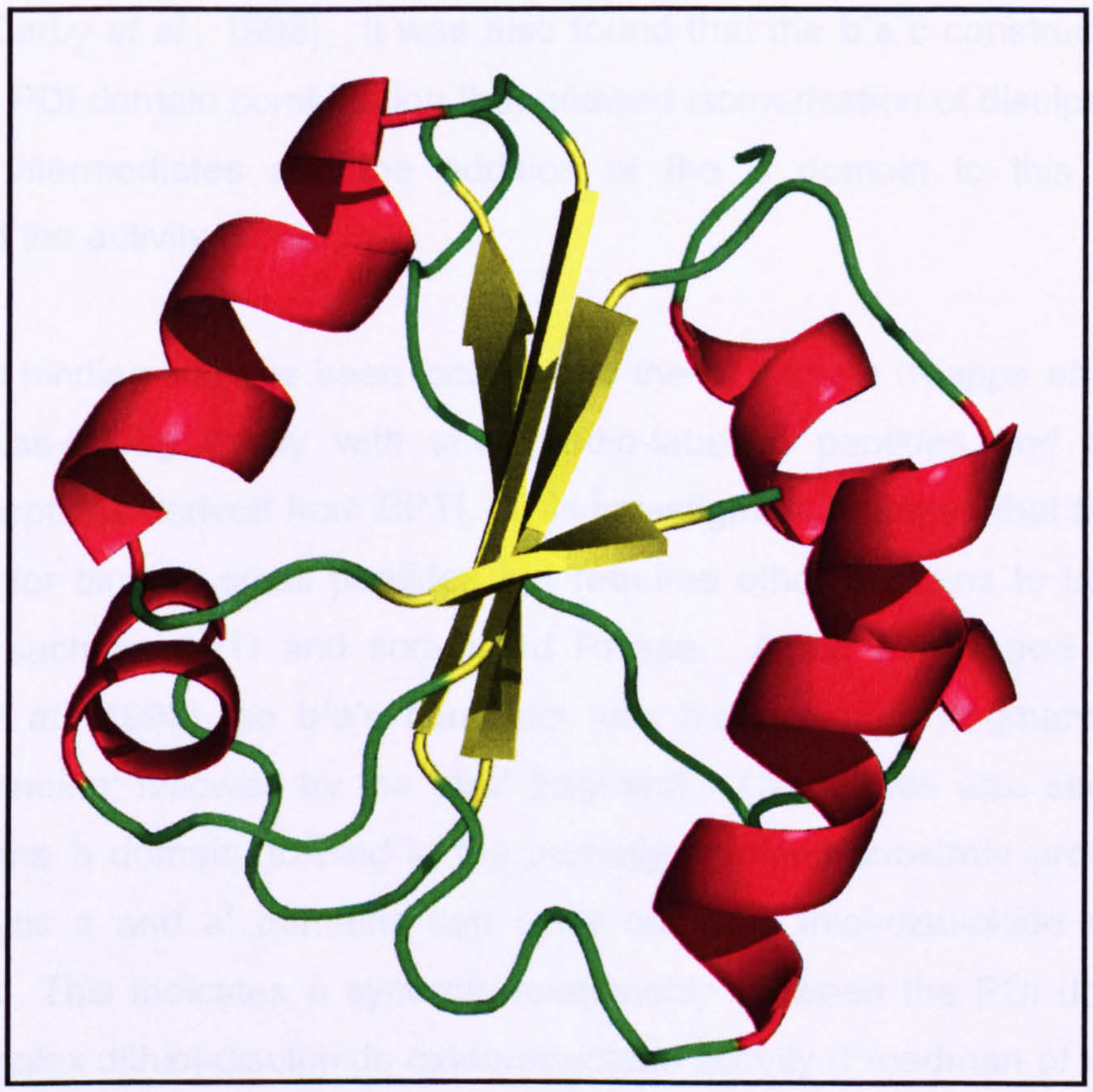

Figure 1-7. Diagram of the human PDI b domain (PDB ID 1BJX). α-helices shown in red and β-sheets shown in yellow.

As previously mentioned, the **a** and **a'** domains alone are not sufficient for complex isomerisation (Darby and Creighton, 1995). Work carried out on various linear combinations of PDI domains using various assays to define transitions between oxidation, reduction and isomerisation showed that only simple thiol-disulphide reactions could be carried out by the **a** and **a'** domains (Darby and Creighton, 1995; Darby *et al.*, 1998). For simple isomerisation reactions the **a** and **a'** domain constructs required the addition of the **b'** domain and complex isomerisation reactions required all four PDI domains but did not require the C-terminal extension. This indicates that the C-terminal extension

appears to play no part in PDI function and the non-catalytic **b** and **b'** domains are functionally important. It was later also shown using bovine pancreatic trypsin inhibitor (BPTI), as a model substrate for PDI to investigate the redox/isomerase activity that the **b'** domain is very important for isomerisation activity (Darby *et al.*, 1998). It was also found that the **b'a'c** construct was the minimum PDI domain combination that allowed isomerisation of disulphide bond in BPTI intermediates and the addition of the **b** domain to this construct increased the activity further.

A specific binding site has been localised to the **b'** domain (Klappa *et al.*, 1998) using cross-linking assay with short radio-labelled peptides and larger 28 residue peptides derived from BPTI. This investigation revealed that the **b'** was essential for binding small peptides but requires other domains to bind larger peptides such as BPTI and scrambled RNase. Again as referred to earlier (Darby *et al.*, 1998) the **b'a'c** construct was the minimum fragment to allow peptide binding; followed by the **abb'** fragment. The **b'** was also shown, with aid from the **b** domain, to bind to the partially unfolded substrate protein while the catalytic **a** and **a'** domains can carry out their thiol-disulphide exchange reactions. This indicates a synergic relationship between the PDI domains to allow complex dithiol-disulphide oxidoreductase activity (Freedman *et al.*, 2002). These results clearly suggest that the **b'** domain is vitally important, holds the principal binding site and is essential for PDI isomerase activity.

1.2.6.4. x-region

The 19 residue **x**-linker/region between domains **b'** and **a'** was first reported ten years ago (Freedman *et al.*, 1998). The **x**-linker is believed to allow flexibility between domains and contains a single tryptophan residue.

1.2.7. PDI function

Mutation studies on the C-terminal cysteine of the C-X-X-C active site, revealed the inability of the mutated PDI to catalyse the refolding of reduced RNase. PDI-Ero1p-mixed disulphides were identified, suggesting that PDI is reoxidised after acting as a protein dithiol-disulphide oxidant (Tu *et al.*, 2000). Similarly mutations in yeast PDIp also indicated PDI was essential for protein disulphide bond formation in the ER.

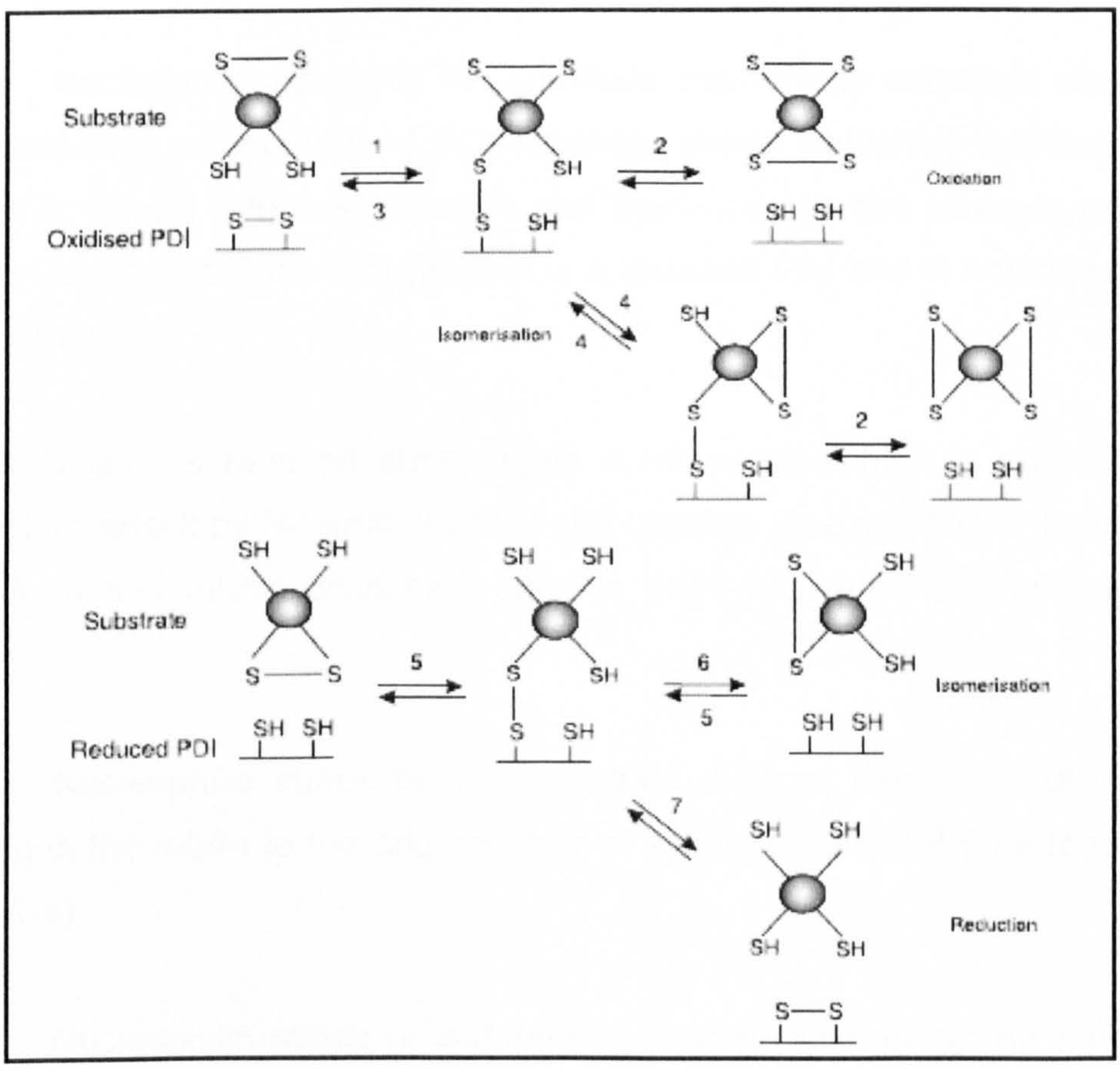

Figure 1-8. Reaction scheme for disulphide oxidation, reduction and isomerisation by PDI and other thioredoxin family members taken from (Hatahet and Ruddock, 2007).

When PDI is in the oxidised state, a mixed disulphide is produced by nucleophilic attack by the substrate thiol on the N-terminal active site cysteine (see reaction 1, Figure 1-8). From this mixed disulphide species there are three possible options:

a) Nucleophilic attack by the substrate thiol on the mixed disulphide species, producing reduced PDI and a net oxidation of the substrate (see reaction 2, Figure 1-8) this reaction prevents PDI getting kinetically trapped as a mixed disulphide species.

b) Nucleophilic attack by the C-terminal cysteine on the mixed disulphide molecule, producing the original products of oxidised PDI and the unchanged substrate (see reaction 3, Figure 1-8).

c) Nucleophilic attack by the substrate thiol on the substrate disulphide bond producing an isomerised PDI-substrate mixed disulphide molecule (see reaction 4, Figure 1-8) this molecule can then undergo the same reaction as stated in a) whereby the final product is a reduced PDI and a net oxidation of the substrate.

If the PDI is in a reduced state, again a mixed disulphide is produced, by nucleophilic attack by N-terminal active site cysteine (see reaction 5, Figure 1-8). Again from this mixed disulphide species there are three possible reaction options:

d) Nucleophilic attack by the substrate thiol on the mixed disulphide, resulting in the return to the original state of substrate and PDI (see reaction 5, Figure 1-8)

e) Nucleophilic attack of a different substrate thiol group on the mixed disulphide species resulting in a net isomerisation of the substrate species (see reaction 6, Figure 1-8)

f) Nucleophilic attack by C-terminal active site cysteine on the mixed disulphide resulting in a net reduction of the substrate species (see reaction 7, Figure 1-8)

PDI kinetics and activity on various model substrates have been well studied to provide information of the *in vitro* mechanisms but fall short of fully explaining *in vivo* reactions (Lyles and Gilbert, 1991; Darby *et al.*, 1994). Mutant forms and various domain combinations have also provided information of the *in vitro* mechanisms, such as substrate reduction or isomerisation (Lyles and Gilbert, 1994; Walker *et al.*, 1996). Mutations of CXXC active site containing domains have been useful as they increase the proportion of on the mixed disulphide species. But there are inherent limitations in these types of analyses, of mixed disulphide complexes undergoing isomerisation or oxidative folding, which would not be trapped with using active site cysteine mutations. However it is possible to obtain data on non-productive kinetically trapped mixed disulphides (Hatahet and Ruddock, 2007). Due to the multiple reaction schemes that take place it is very difficult interpret data obtained in these types of experiments.

The *in vitro* experiments do not fully explain *in vivo* catalysis because the *in vitro* assays are so dependent on conditions. The only *in vivo* analysis for substrate specificity has been to knockout ERp57, which was lethal in embryonic mice (Garbi *et al.*, 2006). However, in ERp57 knock out cells, very few proteins were affected by the loss (Solda *et al.*, 2006), this indicates that ERp57 is involved in the folding of a few proteins or other PDI family proteins are able to compensate for the loss of ERp57.

1.2.8. Molecular chaperone function

PDI has previously been functionally described as a protein folding catalyst, as it increases speed of the rate limiting steps in the protein folding pathway by means of disulphide bond oxidation, reduction and isomerisation. But PDI can also exhibit molecular chaperone activity defined as the ability to prevent the aggregation of substrate molecules and so preventing substrates following a non-productive folding pathway (Wang and Tsou, 1993). PDI was shown to

promote reactivation and prevent aggregation of rhodanese (guanidine-hydrochloride denaturation or thermal inactivation), independent of redox/isomerase function (since rhodanese has no disulphide bonds), *in vitro* and at near stoichiometric concentrations (Song and Wang, 1995). The redox activity was found to be independent from the chaperone activity, since mutated active site residues did not affect the chaperone activity (Winter *et al.*, 2002). Furthermore it was shown that mutations in the principal binding site in **b'** of PDI (Klappa *et al.*, 1998) could reduce the binding and refolding of substrate, but not completely abolish it. This indicates the substrate binding extends through all the PDI domains and further analysis showed that the other domains also contribute to chaperone activity of PDI (Winter *et al.*, 2002). The substrate binding was earlier reported to be weak (Kd>100μM) and was dependent on the length of the substrate backbone binding to PDI, unless cysteine residues are involved, whereby the binding would be much stronger as a result of mixed disulphides forming (Morjana and Gilbert, 1991).

PDI has also been shown to possess chaperone function using a number of other substrate proteins including D-glyceraldehyde-3phosphate dehydrogenase (GAPDH, EC 1.21.12) (Li *et al.*, 2001). GAPDH contains no disulphide bonds and so is used as a target protein to examine the intrinsic chaperone activity devoid of disulphide-linked activity. Recently, the shorter domain combinations of **ab** and **b'a'c** of PDI were shown to lack chaperone function and this was speculated to result from the lack of protein substrate binding requiring more than the peptide-binding site in the **b'** domain (Sun *et al.*, 2000).

1.2.9. PDI as a subunit of collagen prolyl 4-hydroxylase

Collagen prolyl 4-hydroxylase, P4-H (EC 1.14.11.2) is an enzyme which like PDI resides in the lumen of the ER and belongs to a 2-oxoglutarate and non-heme-Fe(II)-dependent oxygenase family of enzymes. P4-H catalyses the formation

of 4-hydroxyproline by hydroxylation of prolines, which requires Fe^{2+}, 2-oxoglutarate, O_2, ascorbate and involves the oxidative decarboxylation of 2-oxoglutarate (Kivirikko *et al.*, 1989). This catalysis is essential in the formation of procollagen polypeptide chains into triple helical molecules. P4-Hs are regarded as suitable drug targets for antifibrotic therapy to combat fibrotic disease which involves an excess build up of connective tissue, primarily collagen.

The collagen P4-H enzyme from vertebrates consists of a $\alpha_2\beta_2$ tetramer with an approximate molecular weights of 240 kDa total (Tuderman *et al.*, 1975), with the α-subunit of 63 kDa (Pankalainen *et al.*, 1970) and the β-subunit of 55 kDa. The β-subunit was found to be PDI (Koivu *et al.*, 1987). It was also later found that two isoforms of the α-subunit existed, α(I) and α(II), which can associate to produce type 1 $\alpha(I)_2\beta_2$ and type 2 $\alpha(II)_2\beta_2$ (Annunen *et al.*, 1997). It was also shown that mixed $\alpha(I)\alpha(II)\beta_2$ did not form in insect coexpression experiments (Annunen *et al.*, 1997). The α(I) subunit consists of 517 residues including a 17 residue signal peptide, the α(II) subunit consists of 514 residues including a 21 residue signal peptide and the two subunits have a 64% sequence identity (Annunen *et al.*, 1997). Recently a third α-subunit was identified, which consisted of 525 residues including a 19 residue signal peptide and has a 36% sequence identity to the α(I) subunit (Van Den Diepstraten *et al.*, 2003). All α-subunits contained five conserved cysteine residues, four of which in the α(I) subunit, form intrachain disulphide bonds, required to maintain the structure and so enable tetramer assembly (Kivirikko and Myllyharju, 1998). The α-subunits were also shown to have an asparagine linked oligosaccharide site, but this was found not to influence catalytic activity or assembly (Lamberg *et al.*, 1995). The catalytic site of P4-H is believed to be within the α-subunit and located in a range of separate sites to allow the binding of all the co-substrates (Kivirikko and Myllyharju, 1998). Using site-directed mutagenesis the Fe^{2+} was located in the α-subunit and the residues His412, Asp414, His483 (Lamberg *et al.*, 1995)

and Lys493 were seen to be involved in 2-oxoglutarate binding (Myllyharju and Kivirikko, 1997).

The β-subunit of P4-H were first identified as PDI in 1987 and it was shown that the PDI in the native human P4-H tetramer has half the isomerase activity of wild type PDI (Koivu *et al.*, 1987). Later it was shown by mutating active site residues in PDI, that PDI activity was not essential for tetramer assembly or P4-H enzyme activity (Vuori *et al.*, 1992a). Also, α-subunits expressed in baculovirus system in insect cells without the β-subunit (PDI) formed highly insoluble and catalytically inactive aggregates (Vuori *et al.*, 1992b). It was demonstrated that P4-H can be assembled in a cell-free system and the solubility of the α-subunit is dependent upon its association with PDI (John *et al.*, 1993). Hence it appears that the β-subunit (PDI) is required to maintain the catalytic activity and ensure the α-subunits of P4-H are in a non-aggregated conformation (Vuori *et al.*, 1992b). The importance of PDI in terms of maintaining the catalytic activity of P4-H was made clear in experiments which showed that co-expression of the α-subunit and BiP, a molecular chaperone, produced soluble but non-active P4-H enzyme. Since expression of α-subunit only resulted in insoluble protein, hence PDI function as a molecular chaperone is required to keep the α-subunit soluble and so ensure P4-H maintains it catalytic activity (Veijola *et al.*, 1996). The β-subunit (PDI) contains the –KDEL ER retention signal, known to interact with receptors. Proteins with this sequence are retrieved from the traffic of secretory proteins and so retained in the lumen of the ER (Munro and Pelham, 1987). Since the α-subunit does not have the ER retention signal, it is believed that the β-subunit is also important to retain P4-H in the ER (Vuori *et al.*, 1992b).

Since there is no full length PDI structure or full P4-H structure, determination of subunit organisation is difficult. Also since the detailed mechanism of PDI substrate binding and function is not known, understanding PDI's role in P4-H is also demanding.

1.2.10. PDI and MTP

Microsomal triglyceride-transfer protein (MTP) was originally isolated from bovine liver microsomes (Wetterau and Zilversmit, 1984) and then later from rat liver and intestines (Wetterau and Zilversmit, 1986). MTP was shown to be composed of two subunits, a 97 kDa α-subunit and the 58 kDa β-subunit (Wetterau and Zilversmit, 1986); the total MTP purified molecular weight was found to be 150 kDa by sedimentation equilibrium experiments (Wetterau *et al.*, 1991), indicating the protein forms a heterodimer of the two protein subunits. The β-subunit was found to be identical to PDI using sequence analysis and immunological characterisation (Wetterau *et al.*, 1990). While no homologous protein was found for α subunit, regions of low sequence homology were found to lipovitellin, which is a protein found in lipoprotein complexes in egg-laying animals (Shoulders *et al.*, 1994).

MTP was initially shown *in vitro* to transfer triglycerides, cholesteryl esters and phospholipids between membranes in the ER (Wetterau and Zilversmit, 1985) and later shown by kinetic analysis that transport occurs by a shuttle mechanism (ping-pong, Bi-Bi kinetics) (Atzel and Wetterau, 1993). It was later shown that MTP plays a vital role in the assembly of lipoproteins, since a rare genetic disease, abetalipoproteinemia, which results in decreased plasma levels of apolipoprotein **b** (apoB) and triglycerides, was caused by a mutation in the MTP gene (Wetterau *et al.*, 1992).

MTP was assumed to be localised to the ER (Wetterau *et al.*, 1997), this was based on the fact that MTP was isolated from the microsomal fraction which consists primarily of the ER membrane. Also, since MTP includes a PDI subunit, which has the –KDEL signal, it was assumed that MTP would be retained in the ER and could not move out into the Golgi. MTP is associated with lipoprotein

assembly, but plasma very low density lipoproteins (VLDL) are believed to be assembled by sequential addition of lipids to already existing nascent molecules in the ER and Golgi (Higgins and Hutson, 1984). Hence contrary to the MTP ER localisation hypotheses, it was speculated that MTP was not restricted to the ER. Later it was found using immunohistochemical studies and confocal microscopy, that MTP was indeed co-localised to the ER and to within the Golgi apparatus (Swift *et al.*, 2003).

MTP was shown to possess two classes of lipid binding sites; the first class, rapidly transports neutral lipids and phospholipids between MTP and membranes. The second class, slowly transports lipid molecules between MTP and membranes and appears to be selective for phospholipids (Atzel and Wetterau, 1994). The lipid binding pocket was speculated to be hydrophobic and flexible, since it can bind neutral and polar lipid molecules; the influence on substrate binding from the β-subunit (PDI) binding site is not known.

The carboxy-terminal 30 amino acids of MTP α-subunit was found to be vital in interactions between the α and β subunits, since truncated α-subunit was unable to form a soluble complex with the β-subunit (PDI) (Ricci *et al.*, 1995). The PDI β-subunit is known to be vital to ensuring the α-subunit maintains a catalytically active state and soluble complex, since dissociated α-subunits formed inactive insoluble protein aggregates (Wetterau *et al.*, 1991b); this was further proven, since individually expressed α-subunit also formed inactive, insoluble aggregates (Lamberg *et al.*, 1996). Also shown was that the redox activity of PDI was not essential to ensure assembly and maintain functional activity of MTP. Both active sites in PDI were mutated and yet the comparable MTP activity and solubility was recorded from both mutant and wild-type co-expression experiments with PDI (Lamberg *et al.*, 1996).

1.3. PDI Family

PDI is the archetypal representative of the family, proteins in the family comprise of at least one thioredoxin-like domain. This section aims to summarise the domain organisation and highlight the structural/functional information available for human PDI family member domains and for PDI-like proteins from other species as shown in Table 1.

1.3.1. Human PDI family members and structurally determined related proteins

Although the name protein disulphide isomerase implies isomerase ability; this is not true or has not been experimentally proven for all family members, such as PDILT (van Lith *et al.*, 2007). In the case of ERp27 and ERp28, it is likely that since they lack any catalytic domains, they show no isomerase activity. Hence, all family members do not share a common function, they are related and grouped on the basis of their common structure (in terms of the thioredoxin fold which can be catalytic or non-catalytic), sequence similarities and ER localisation. The three related PDI proteins from *Saccharomyces cerevisiae, Oryctolagus cuniculus* and *Humicola insolens,* are included here since they are the only related proteins which share an appreciable sequence similarity coinciding with available three-dimensional structural data.

All family members share in common at least one domain either **a** or **b**. ERp28/29 (named either in literature) is an exceptional member in that it has a non-thioredoxin like, helical D-domain at the C-terminus (Liepinsh *et al.*, 2001). A second exception is ERdj5 (Cunnea *et al.*, 2003) which has a N-terminal J-domain, which associates with BiP (Hsp70 chaperone family member) in an

ATP-dependant manner which enhances ATPase activity (Hosoda *et al.*, 2003). Also, most PDI family members are ER localised, with the aid of C-terminal –KDEL or a –KDEL variant ER retention signal. The exceptions to this are Hag2 (KTEL), Hag3 (QSEL), ERp18 (EDEL), which may not be localised to the ER but this is not proven since there is no *H. sapiens* expression analysis to date (Persson *et al.*, 2005). ERp44 was initially identified as a ER-located binding partner for Ero1, involved in thiol-mediated retention of client proteins (Anelli *et al.*, 2002). ERp44 has a –RDEL motif but was found to be not retained in the ER, but co-localised to the ER and ER-Golgi intermediate compartment and *cis*-Golgi (Gilchrist *et al.*, 2006).

Table 1:

Species Source	Amino Acids	Protein	Domain architecture	PDB code
Homo sapiens	508	PDI	a-b-b'-x-a'-c	a : 1MEK b : 1BJX a' : 1X5C
Homo sapiens	505	ERp57	a-b-b'-x-a'-c	a : 2ALB bb': 2H8L a':2DMM
Homo sapiens	645	ERp72	a-a-b-b'-x-a'-x	a : 2DJ1 a : 2DJ2 a' : 2DJ3
Homo sapiens	172	ERp18	a	a :1SEN
Homo sapiens	261	ERp28/ERp29	b-D	b-D: 1OVN
Homo sapiens	440	P5	a-a'-b	a : 2DML a' : 1X5D
Homo sapiens	432	ERp46/endoPDI	a-a-a'	a' : 2DIZ
Homo sapiens	525	PDIp	a-b-b'-x-a'-c	
Homo sapiens	519	PDIr	b-a-a-a-c	
Homo sapiens	793	ERdj5	J-a-b-a-a-a'	
Homo sapiens	273	ERp27	b-b'	
Homo sapiens	584	PDILT	a-b-b'-x-a'-c	
Homo sapiens	406	ERp44	a-b-b'	
Homo sapiens	165	HAG3	a	
Homo sapiens	175	HAG2	a	
Saccharomyces cerevisiae	522	PDI1p	a-b-b'-x-a'-c	2B5E
Oryctolagus cuniculus	367	Calsequestrin	I-II-III	1A8Y
Humicola insolens	133	PDI	b'	2DJK

Table 1. Summary of protein information for 15 human ER resident-PDI-family members and three PDI-like proteins from other species. Highlighted in red are domains which have solved three-dimensional structures. J- dnaJ domain, D- helical domain.

1.3.2. Structural/functional data available for PDI family and related protein

The most complete structural information for a human PDI family member is that for ERp57, whereby the individual **a** and **a'** domains have been structurally determined by NMR (Silvennoinen *et al.*, 2001; Silvennoinen *et al.*, 2005) and the **bb'** domain combination by x-ray crystallography (Kozlov *et al.*, 2006). No structural information for a full length human PDI family member, incorporating all four **a**, **b**, **b'** and **a'** domain is available and apart from the **bb'** domain combination of ERp57, no other human **b'** domain structural information is available. Until 2006 the closest protein with sequence similarity and three-dimensional structural data was calsequestrin which was determined in 1998 by X-ray crystallography (Wang *et al.*, 1998). It was not until 2006, that the **bb'** domain combination was structurally determined by x-ray crystallography (Kozlov *et al.*, 2006), shortly after the NMR assignments of the thermophilic fungal *Humicola insolens* **b'** domain were published (Nakano *et al.*, 2006). Also that same year and most importantly, the first full length PDI structure from *Saccharomyces cerevisiae* (Yeast) was reported, solved by x-ray crystallography (Tian *et al.*, 2006).

No three-dimensional structure of the full length human PDI molecule has been solved to date or of any other human catalytically active PDI family member. The only experimental data available on full length human PDI comes, again very recently, from low resolution small angle x-ray scattering (SAXS) (Li *et al.*, 2006). As shown in Figure 1-9, the SAXS data revealed an elliptical cylinder structure which when modelled with two **a** domains (replacing **a'** with **a**, since no high resolution **a'** structure is available) and two **b** domains (replacing **b'** with **b**, since no **b'** domain structure is available) appear to show PDI having an annular arrangement. Interestingly, this is similar to the twisted "U" structure of the full

length *Saccharomyces cerevisiae* PDI (Tian *et al.*, 2006), discussed in more detail in section 1.3.5. Although as a result of the low resolution of the SAXS data, no definitive comparisons can be made.

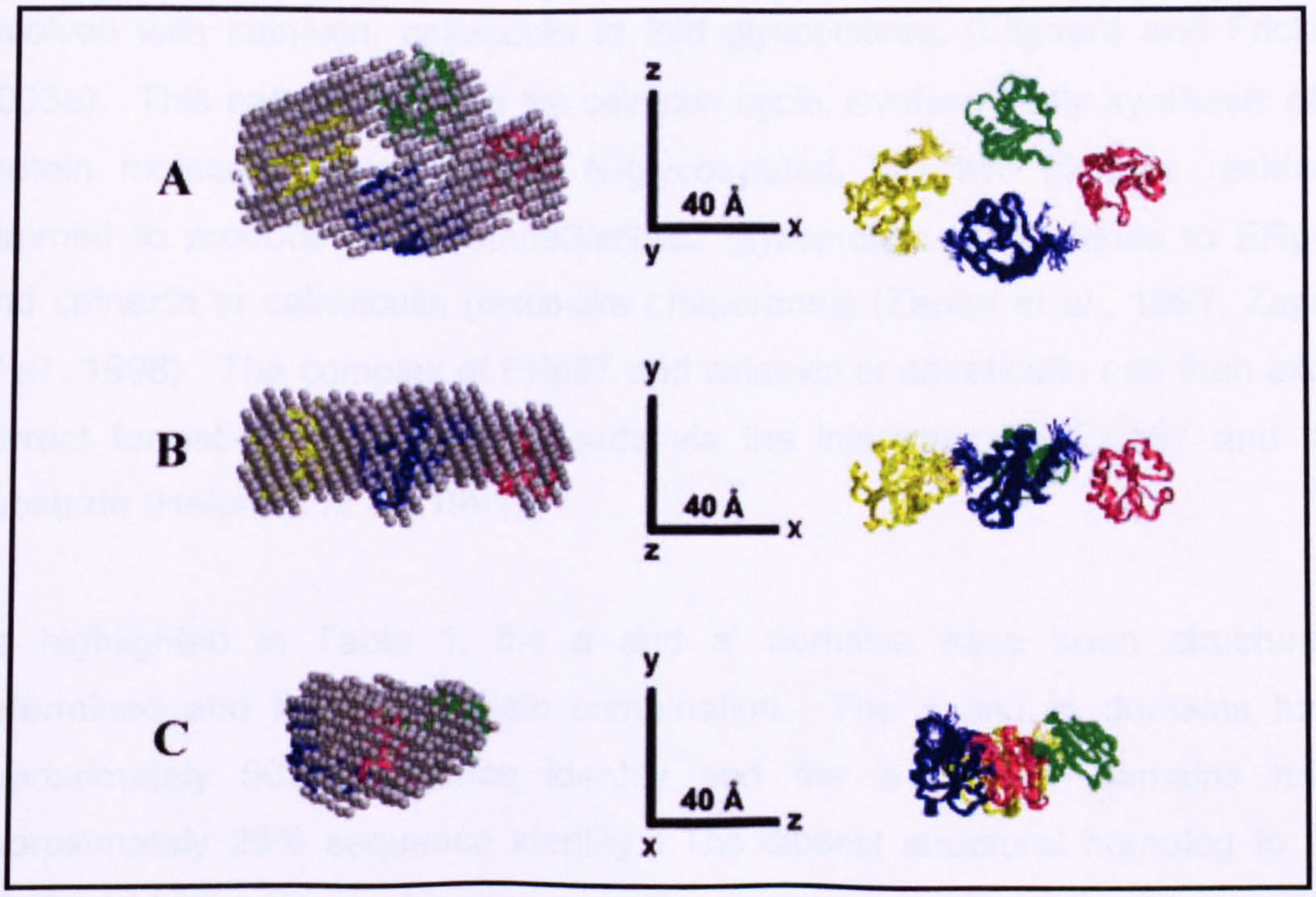

Figure 1-9. Model of human PDI structure using SAXS data, superimposed with two b domains, 1BJX (pink and green domains) and two a domains, 1MEK (blue and yellow), showing an annular arrangement of a-b-b'-a' in the counter-clockwise direction taken from Li *et al*, 2006.

1.3.3. ERp57

Other than PDI, ERp57 is the most characterised of all the PDI family members. ERp57 has 505 residues including a 24 amino-acid N-terminal signal sequence (Bourdi *et al.*, 1995) and a –QDEL ER retrieval sequence. ERp57 has also been shown to have the same domain architecture as PDI, in that it has two thioredoxin-like catalytic active **a** and **a'** domains with similar and similarly positioned active site residues (Ferrari and Soling, 1999). The sequence identity between PDI and ERp57 has been reported to be 29% and 56% similarity, but ERp57 has been shown to be unable to substitute for PDI as the β-subunit in P-4H (Koivunen *et al.*, 1996).

ERp57 is multifunctional, like PDI and has been shown, *in vitro*, to act as a oxidoreductase (Frickel *et al.*, 2004). ERp57 has also been shown to be involved with calnexin, calreticulin to fold glycoproteins (Ellgaard and Frickel, 2003a). This pathway termed the calnexin cycle, involves firstly synthesis of a protein molecule, which is then N-glycosylated, has two glucose residues trimmed to produce a Glc1Man9GlcNAc2 glycoprotein which binds to ERp57 and calnexin or calreticulin (lectin-like chaperones) (Zapun *et al.*, 1997; Zapun *et al.*, 1998). The complex of ERp57 and calnexin or calreticulin can then allow correct formation of disulphide bonds via the interaction of ERp57 and the substrate (Helenius *et al.*, 1997).

As highlighted in Table 1, the **a** and **a'** domains have been structurally determined and the **bb'** domain combination. The **a** and **a'** domains have approximately 50% sequence identity and the **b** and **b'** domains have approximately 20% sequence identity. The closest structural homolog to the ERp57 **bb'** domain combination was *Saccharomyces cerevisiae* (Yeast) PDI, which when superimposed gave a rmsd (root mean squared deviation) of 3.5 Angstroms over 199 residues, indicating a similar orientation of both domains in both proteins (Kozlov *et al.*, 2006). This **b'** domain from ERp57 is the only PDI human family member **b'** domain structure available, highlighting the difficulty in crystallisation and obtaining NMR spectral data on this domain.

The **bb'** ERp57 construct (PDB code 2H8I) includes the **x**-region; when the **b'** sequence of the **bb'** construct is aligned with human PDI **b'x**, a 22% sequence identity and a 39% sequence similarity is observed. Whereas if human **b'** domain alone is aligned with the **b'** region of **bb'** ERp57 domain then 18% sequence identity and a 32% sequence similarity is observed. The human **b'** domain consists of 123 residues, human **b'x** has 147 residues, the ERp57 **b'** sequence has 139; too long to be a lone domain and since the sequence similarity and identity increase when aligned with the human **b'x** construct, it

appears that the ERp57 **bb'** di-domain construct indeed includes the **x**-region. Interestingly, as shown in Figure 1-10, the C-terminal section or **x**-region of construct appears free and not bound to or in close proximity to the rest of the **b'** domain, this observation is discussed in greater detail in the discussion.

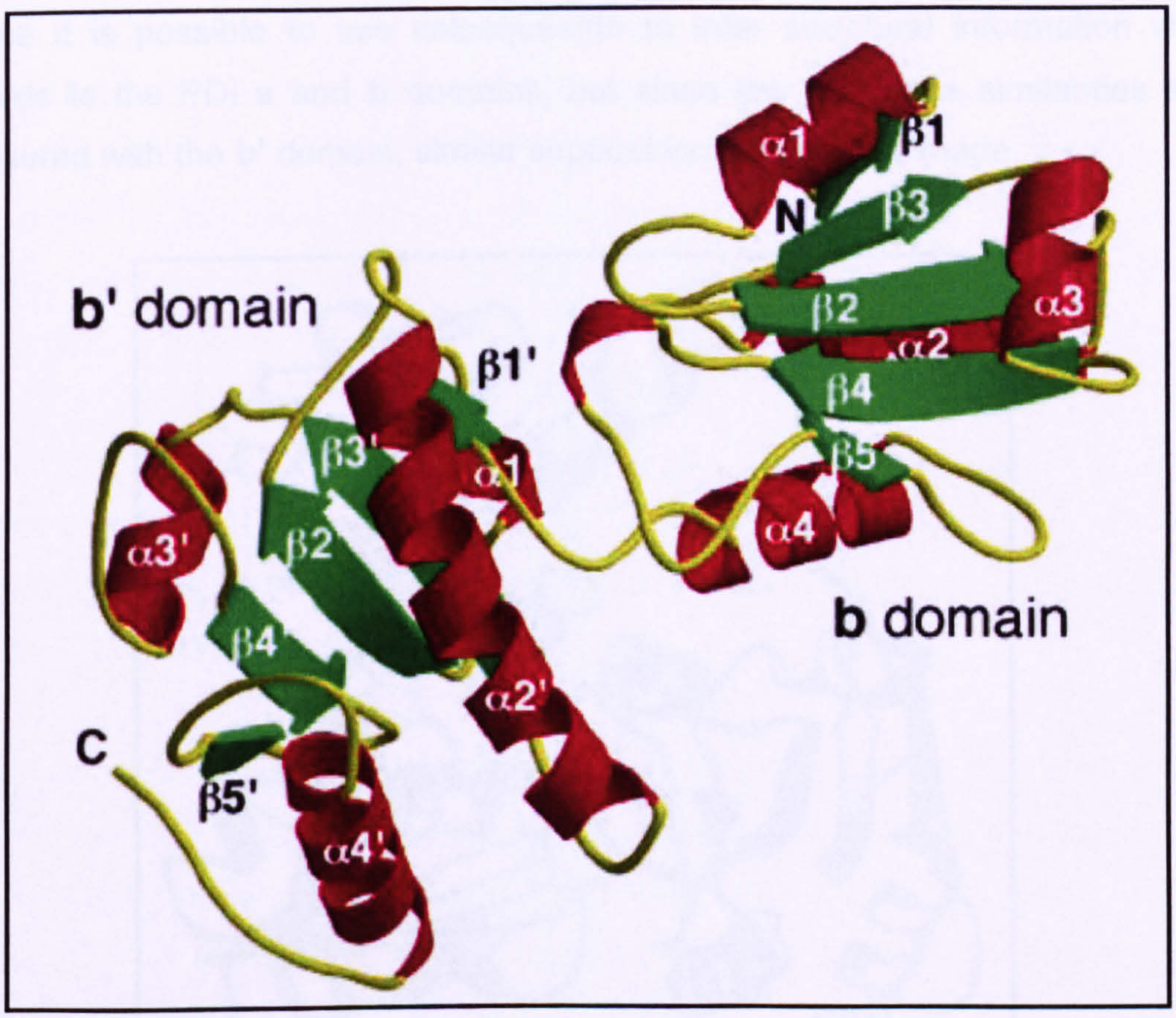

Figure 1-10. Secondary structure diagram of b, b' domains and x-region at the C-terminus of ERp57 taken from Koszlov *et al*, 2006.

1.3.4. Calsequestrin

Calsequestrin is a 40 kDa protein, located in the luminal spaces of the terminal cisternae of the sarcoplasmic reticulum in muscle cells and is involved in the regulation of the Ca^{2+} ion channel (Kawasaki and Kasai, 1994). Until as recently as 2006, the closest protein with sequence similarity and three-dimensional structural data to PDI was calsequestrin, which was solved in 1998.

Calsequestrin is made up of three domains (DI, DII, DIII) all thioredoxin-like but showing low sequence similarity to thioredoxin (Wang *et al.*, 1998). The highest similarity between individual calsequestrin domains and individual PDI domains were 39% similarity for DI and PDI **a** domain and 46% for DII and PDI **b** domain, but the DIII gave very low similarity to all PDI domains, 6.7% with PDI **b'** domain. Hence it is possible to use calsequestrin to infer structural information with regards to the PDI **a** and **b** domains, but since low sequence similarities are measured with the **b'** domain, similar suppositions can not be made.

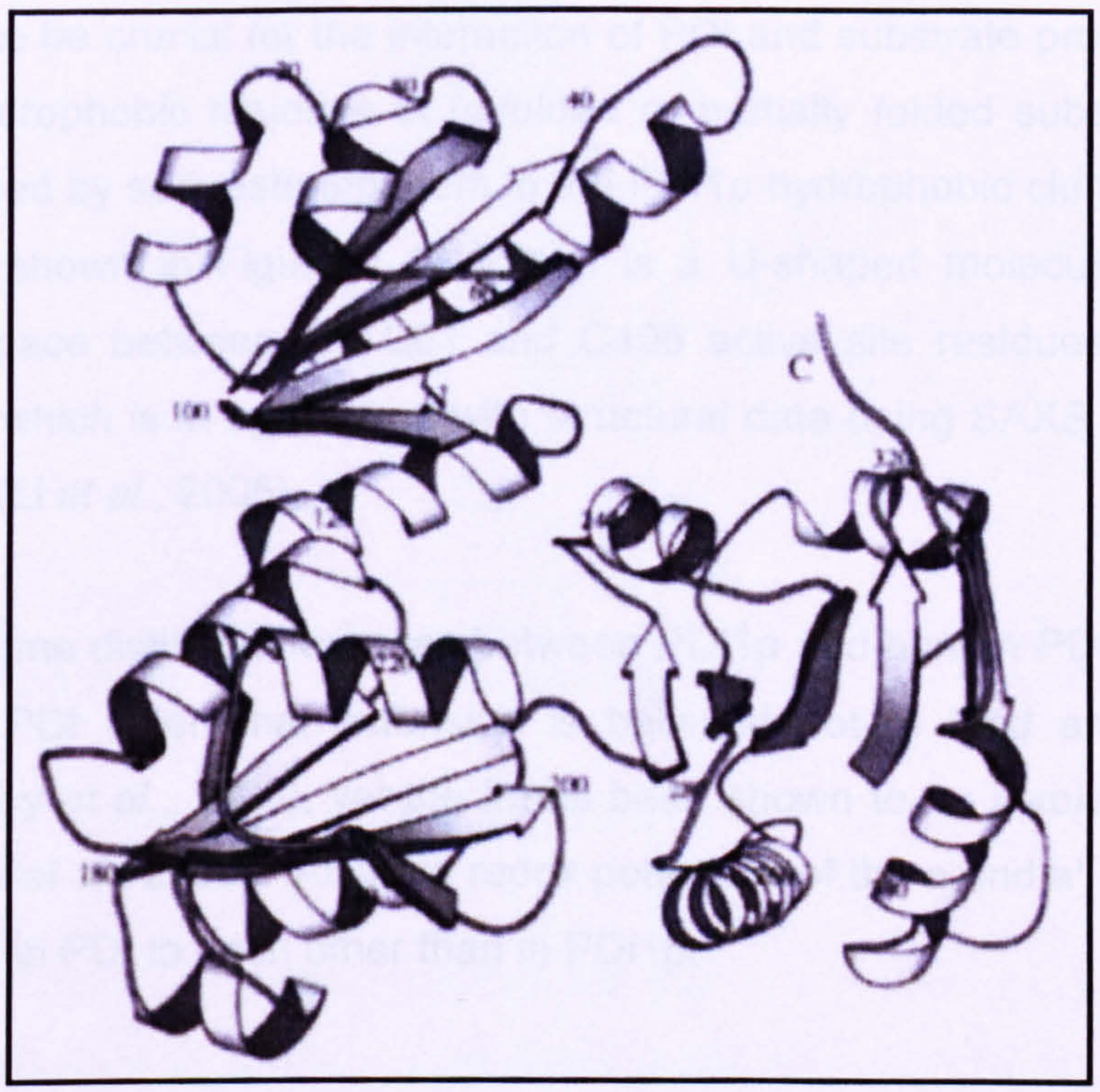

Figure 1-11. Secondary structure diagram of the Calsequestrin taken from Wang et al, 1998.

1.3.5. Yeast – PDI1p

PDI1p from *Saccharomyces cerevisiae* (Yeast) was the first full length PDI structure solved by x-ray crystallography in 2006 at 2.4 Angstrom resolution (Tian *et al.*, 2006). The sequence similarity between PDI and PDI1p is high and

by superimposing the **a** and **b** domains of PDI1p and PDI, there were rmsd's of 1.4 Angstroms for 106 Cα atoms and 1.8 Angstroms for 75 Cα atoms, respectively. There is an insertion in PDI1p, in the middle of the **b'** domain, the implication of which is as yet unknown. There is also reported to be an approximately 700 Angstrom^2 buried surface area between the **b** and **b'** domains, this interface poses as a hydrophobic patch and is surrounded by the active sites from the **a** and **a'** domain, producing a continuous hydrophobic surface. This hydrophobic cleft is speculated to be large enough to accommodate a folded protein of approximately 100 residues. This is speculated to be crucial for the interaction of PDI and substrate proteins. Since exposed hydrophobic residues in unfolded or partially folded substrates could be suppressed by sequestering them in the PDI1p hydrophobic cleft (Tian *et al.*, 2006). As shown in Figure 1-12 PDI1p is a U-shaped molecule with a 28 Angstrom space between the C61 and C406 active site residues which face each other, which is in agreement with structural data using SAXS collected on human PDI (Li *et al.*, 2006).

There are some distinct differences between PDI1p and human PDI, one is that the human PDI C-terminal extension is believed not to hold any functional activity (Darby *et al.*, 1998), yet the it has been shown to be involved in PDI1p activity (Tian *et al.*, 2006). Also the redox potentials of the **a** and **a'** domains are more similar in PDI to each other than in PDI1p.

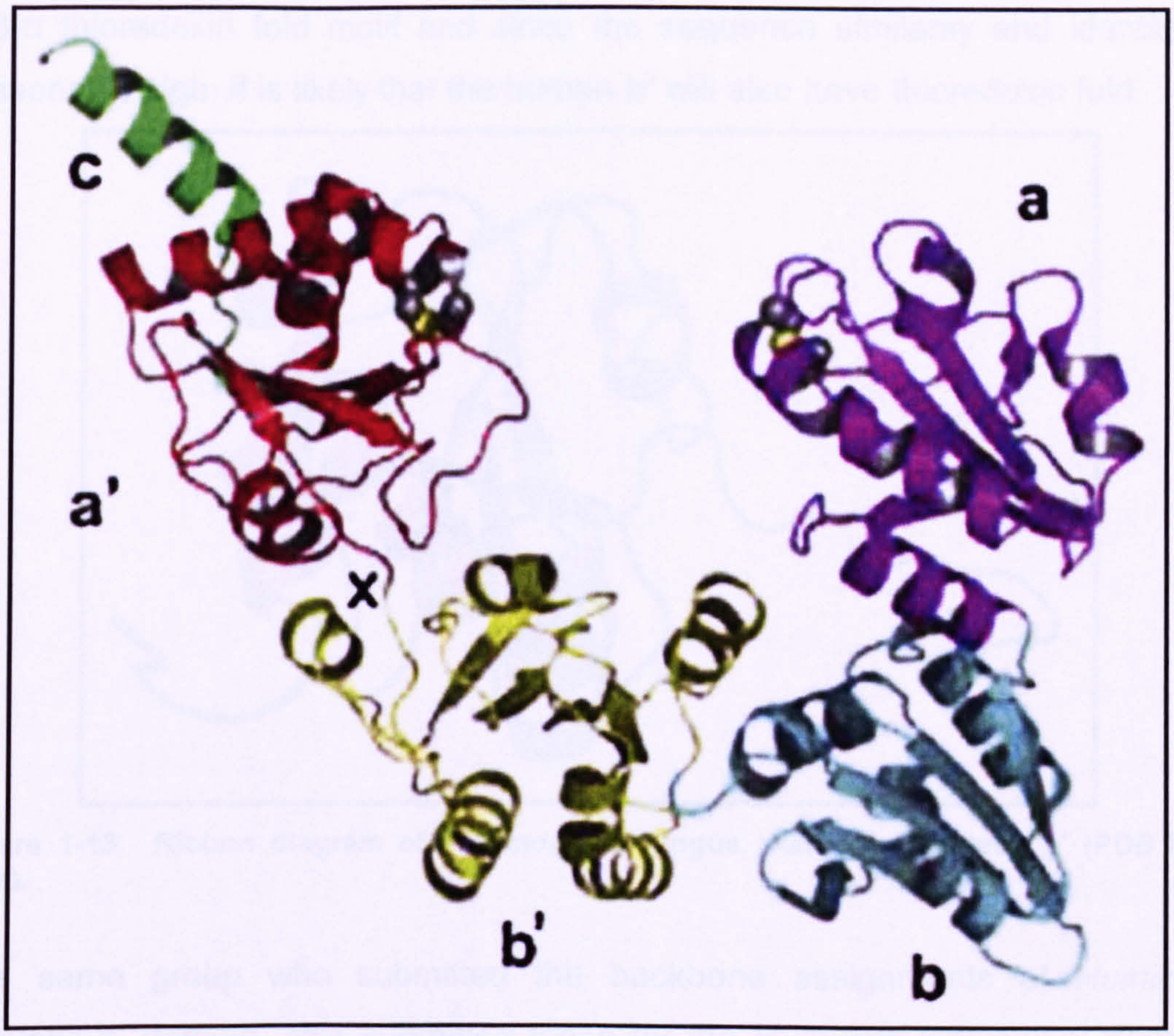

Figure 1-12. Ribbon diagram of PDI1p domains. The active site cysteines in the a and a' domain shown in space fill representation and sulphur atoms in yellow, taken from Tian *et al* 2006.

1.3.6. Thermophilic fungus b' (2djk)

No functional data has been reported for the PDI protein isolated from the thermophilic fungus, *Humicola insolens,* but a low resolution NMR structure has been deposited, PDB code 2djk (Nakano *et al.*, to be published). The reported **b'** domain isolated from *Humicola insolens* consists of 133 residues, longer than the human **b'** domain and a greater sequence similarity is recorded when compared to **b'x** than the **b'** domain. Hence it is likely that the *Humicola insolens* **b'** domain (2djk) is actually the **b'x** domain combination. The human **b'x** and 2djk has a 26% sequence identity and has sequence similarity of 43%. The secondary structure of 2djk does appear to have the consensus α-β-α-β-α-

β-β-α thioredoxin fold motif and since the sequence similarity and identity is reasonably high, it is likely that the human **b'** will also have thioredoxin fold.

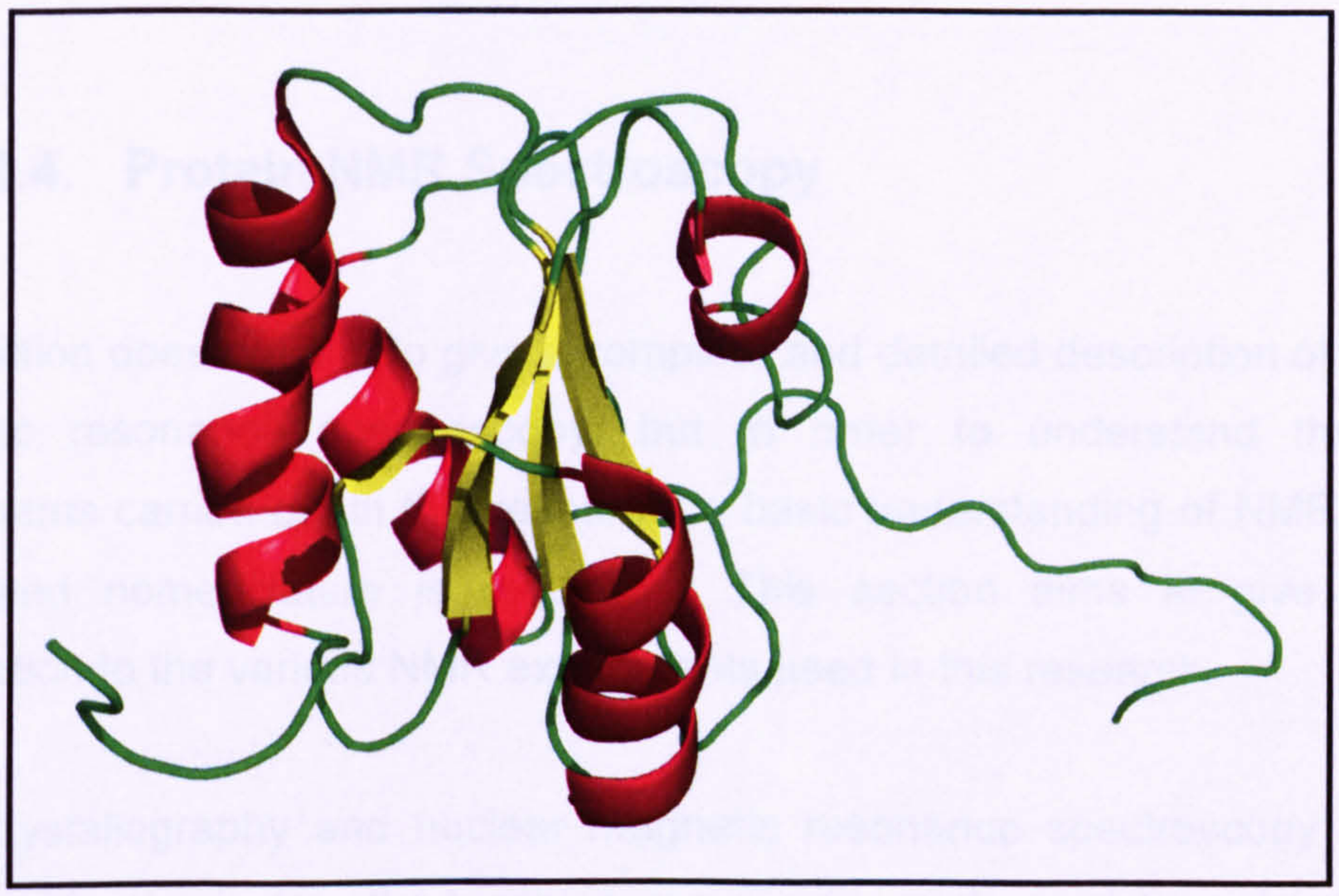

Figure 1-13. Ribbon diagram of Thermophilic fungus Humicola insolens b' (PDB ID - 2djk).

The same group who submitted the backbone assignments of *Humicola insolens* **b'** domain, also submitted those for the *Humicola insolens* **a'** domain (Nakano *et al.*, 2006). But as yet no full structure has been deposited in the PDB data bank.

1.3.7. Human bb' chemical shifts

The most important and relatable structural information to date was the **bb'** domain NMR chemical shifts assignments, published two years into this study in 2006 (Denisov *et al.*, 2006). The chemical assignments are from human PDI residues 135-357, this construct has an extra 5 residues on the N-terminus which are not found in the human PDI sequence. But more importantly, the construct has only the first 7 residues of the widely accepted **x**-region (Freedman *et al.*, 1998) and lacks the final 12 residues of the **x**-region. The

implications of this trimmed version of essentially the **bb'x** domain are discussed later in the Discussion section.

1.4. Protein NMR Spectroscopy

This section does not aim to give a complete and detailed description of nuclear magnetic resonance spectroscopy, but in order to understand the NMR experiments carried out in this research, a basic understanding of NMR theory, terms and nomenclature is required. This section aims to give a brief introduction to the various NMR experiments used in this research.

X-ray crystallography and nuclear magnetic resonance spectroscopy are the only two bio-physical methods which can provide high resolution structures of biological molecules at atomic resolution. The advantages of NMR over crystallography is that NMR uses molecules in solution and so crystallisation is not required; hence NMR is not limited to those proteins that crystallise well. Without the need for crystallisation NMR samples are not prone to crystallisation packing affects, so NMR will provide information on a protein structure which is closer to its native-like structure in the cell. Furthermore, crystallisation is often a bottleneck in structural determination in x-ray crystallography, whereas NMR is more reliable and the major limiting factor is the time taken to analyse data.

NMR also allows the study of molecular properties in addition to structural data, such as protein flexibility and ligand binding. This is not possible with crystallography as it would require a new ligand bound crystal. NMR can provide information about conformation or chemical exchange, internal mobility and dynamics at timescales ranging from picoseconds to seconds. Hence, flexible loops regions can easily be seen by NMR but in crystallography these are not visible due to spatial averaging of the electron density. The greatest advantage of NMR over crystallography is the analysis of interactions with other

molecules such as protein-protein and protein-ligand. For example a protein ligand can be titrated into a protein sample, which induces chemical shift changes in the spectrum for residues which lie near the binding site. Therefore chemical shift mapping can locate the binding site and determine the dissociation constants (K_D).

The only drawback is an upper weight limit of proteins for NMR structural determination, which is approximately 50kDa, but with recent advances in magnetic field strengths, the use of segmental labelling (Otomo *et al.*, 1999) and cryoprobes NMR spectra can be collected for molecules over 100kDa. But, x-ray crystallography provides the only method of high-resolution structural determination over these NMR limits.

In the case of PDI and PDI family members, protein flexibility is likely to be the main reason why crystallisation of the **b'** domain have failed, this is an inherent problem in crystallisation of proteins, it is widely known that compact, stable proteins produce the best diffracting crystals for study. Until very recently structural data on PDI family members was derived from NMR studies only, as discussed in the previous section. Therefore, NMR offers the most optimistic method to obtain structural data on the **b'** domain.

1.4.1. Basic principles of NMR

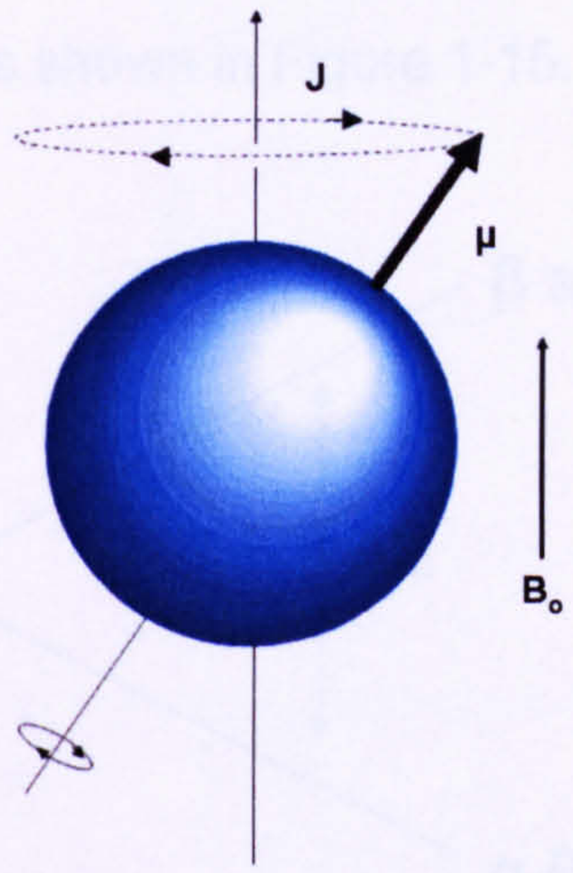

Figure 1-14. Diagram representing the motion of a nucleus in an external magnetic field, adapted from (Evans, 1995). B_o – magnetic field, μ – magnetic moment and J – nuclear angular momentum.

Nuclear magnetic resonance relies on atoms which have a magnetic moment or spin I>0. Nuclei with an even mass number and even charge number have no nuclear spin (I = 0). Nuclei with an odd mass have a half integral spin (I=$^1/_2$, $^3/_2$ etc.). Nuclei with an even mass and odd charge number have an integral spin (I = 1, 2 etc.). The most common biological nuclei ^{12}C, ^{14}N, ^{16}O have an I =0 and so do not give NMR spectra. Fortunately ^1H has a nuclear spin of $^1/_2$. Carbon and Nitrogen nuclei can only be observed when isotopically labelled with ^{13}C and ^{15}N which have an I = $^1/_2$, to enable the collection of NMR spectra. The nuclear magnetic moment (μ) is given by $\mu = \gamma I$. The gyromagnetic ratio, γ, is a constant which determines the resonant frequency of the nucleus in a given field. A nucleus with a nuclear spin of $^1/_2$ can be viewed as a small bar magnet (as shown in Figure 1-14), where a nucleus in an applied magnetic field B_o, the magnetic moment μ experiences a torque which is the vector product of the nuclear angular momentum, J. The nucleus, when placed in a static external field, has an energy which varies depending on the orientation to the field. The

two possible energies are quantised as α or β corresponding to parallel or antiparallel orientations of the nucleus to the external field. Hence, when a NMR sample is placed in a static magnetic field B_o the spins of molecules are either in an α or β state as shown in Figure 1-15.

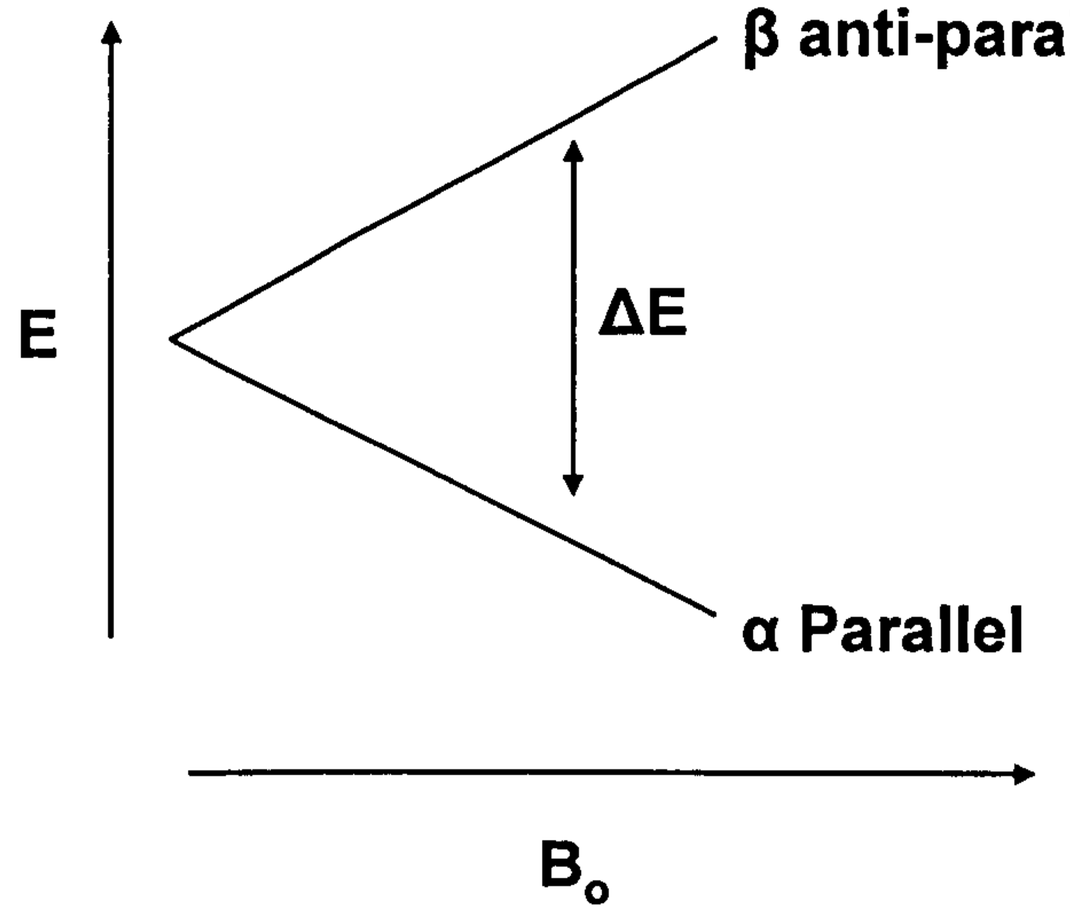

Figure 1-15. Diagram showing the energy levels for a nucleus. B_o - Magnetic field, E - Energy, ΔE - Change in energy.

The population between α and β energy states will be unequal, since a parallel orientation on the z axis, which is the same direction as the magnetic field has a lower energy than the antiparallel one. Hence there is a net magnetisation in the z axis (called M_o) and so ensembles of spins are aligned to the z axis as shown in Figure 1-16.

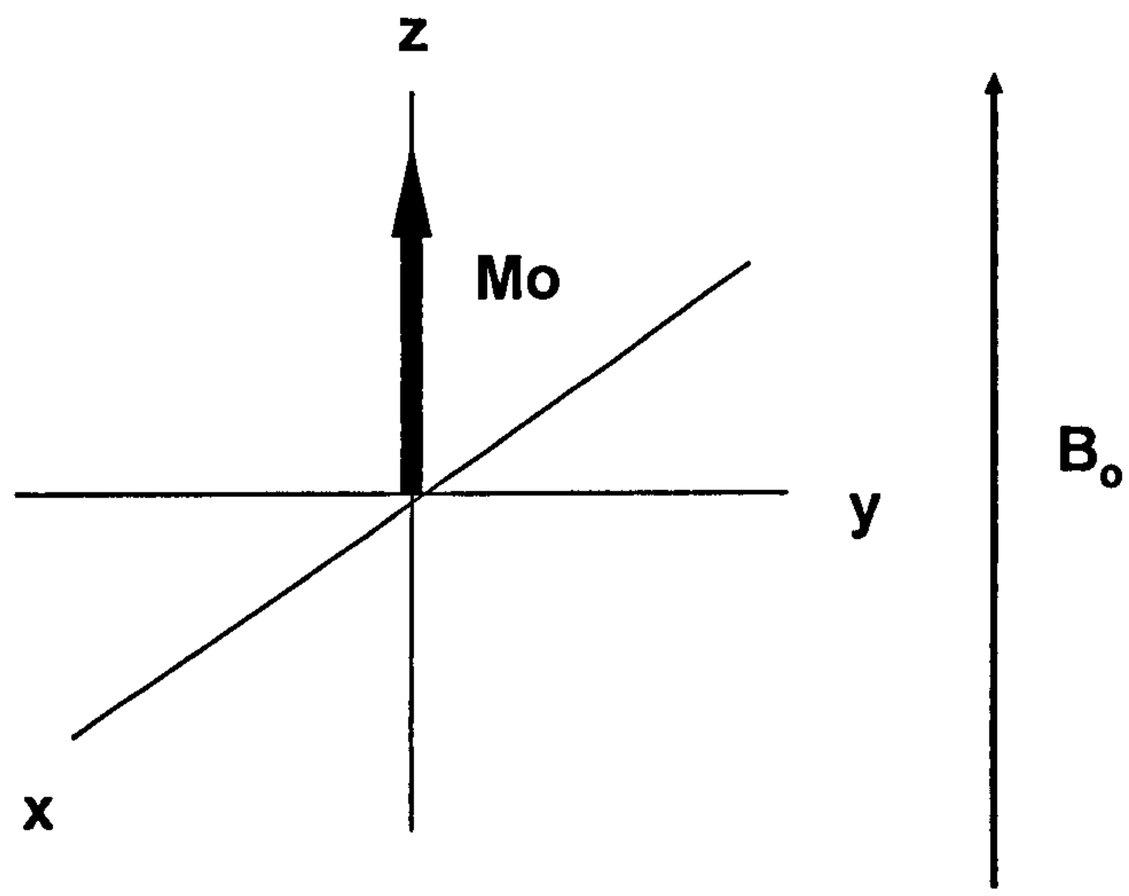

Figure 1-16. Diagram showing the z axis magnetisation (M_o) in the same direction as the magnetic field (B_o)

So in order to induce nuclear magnetic resonance, an oscillatory magnetic field is applied at a frequency which matches the difference between the two spin energies (ΔE). These resonance frequencies correspond to radiofrequency wavelengths at the low frequency end of the electromagnetic spectrum. The frequencies required to induce resonance of all the nuclei in a protein will vary slightly as a result of chemical environment, so a range of frequencies is required. The range of frequencies can be encoded into a short radiofrequency pulse which is then applied, allowing an induction for the complete frequency spectrum in one experiment called a Fourier transform NMR. The output of this experiment is a super-position of the frequencies for every spin in protein molecule as a function of time F(t), this is mathematically transformed into F(w) an intensity as a function of frequency.

1.4.2. Chemical shift

The resonant frequencies also called chemical shifts, measured in *parts per million* (ppm) and vary in different nuclei as a result of shielding of magnetic field by electrons in the local environment of the nucleus. Chemical shift (δ) is used to describe the position of a frequency, V, of a NMR signal in relation to the

signal from a standard of known frequency, as defined by $\delta = ((V_{ref} - V_{obs})/V_{ref})x\ 10^6$ where V_{ref} is the resonance frequency of a standard substance, V_{obs} is the resonance frequency observed (Hammes, 2005). Chemical shifts are dimensionless and are independent of magnetic field strength.

So it is possible to distinguish between various nuclei spins, for example the shielding from a backbone amide proton HN will give a different chemical shift to the Hα in the same protein because each hydrogen nucleus has a different chemical environment. In this way, NMR can be used to assign individual NMR signals to specific atoms in a molecule.

1.4.3. HSQC

Heteronuclear Single Quantum Correlation (HSQC) is a 2D experiment which detects correlations between the ^1H atoms and the directly bonded NMR active heteroatoms ^{15}N or ^{13}C (Maudsley and Ernst, 1977). NMR samples prepared for HSQC analysis are ^{15}N-labelled and so provide information for the amide backbone and side chain NH_2 groups (seen as a pair of peaks in the ^1H dimension, as shown in Figure 1-17).

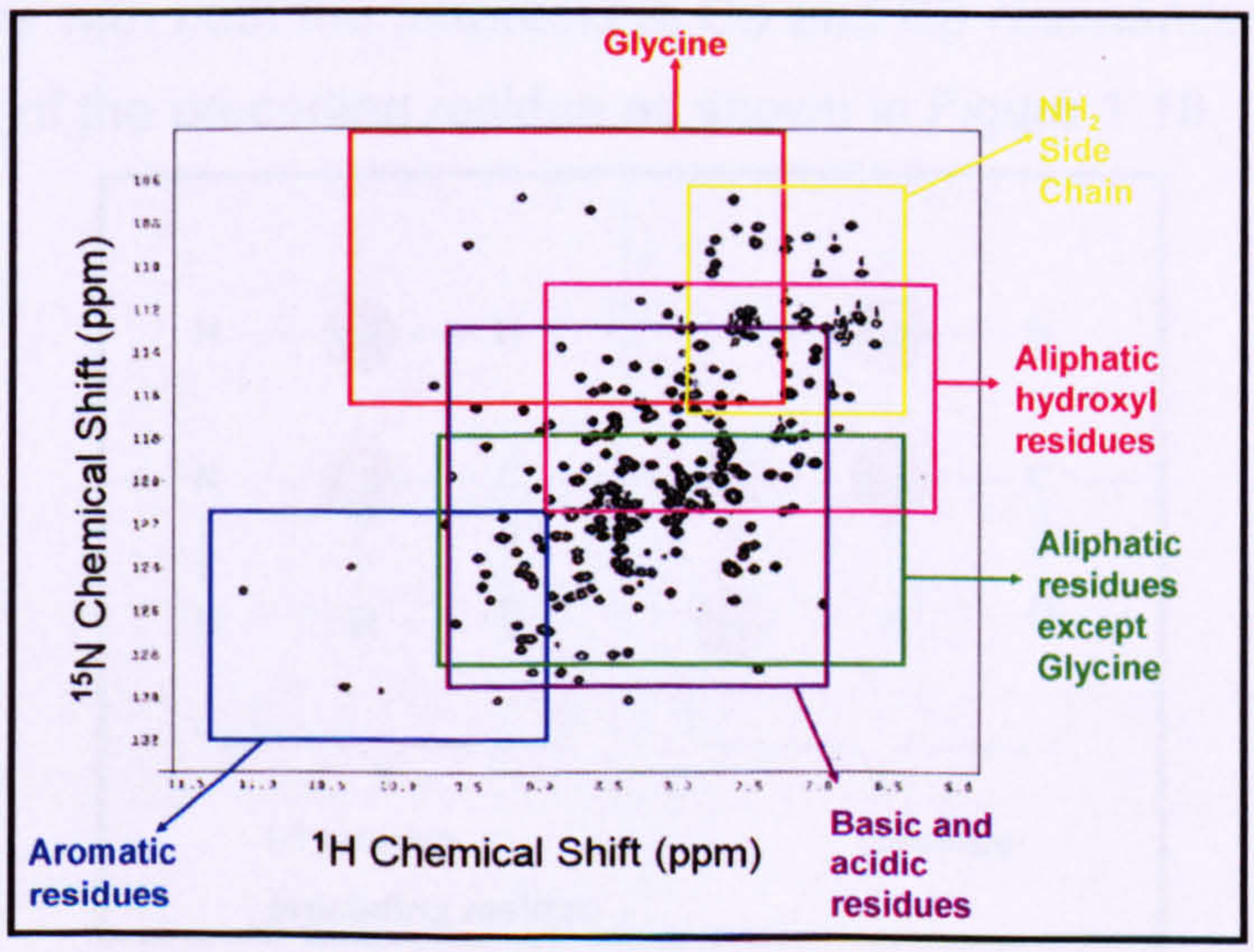

Figure 1-17. $^{15}N/^1H$ HSQC spectrum of ^{15}N-bb'x domain at 40°C used to highlight the average chemical shifts for residues groups.

The chemical shifts highlighted in Figure 1-17 are average chemical shifts for different protein side chains groups and glycine; actual recorded chemical shifts may vary since they are dependent on the local chemical environment.

1.4.4. Through-bond experiments for sequential assignment

In order to order to obtain structural data on macromolecules, multidimensional NMR is required. Multidimensional NMR relies on the principle of linking two resonances, either through space or through a small number of chemical bonds. The resulting cross-peaks give information on spatial relationships. The method of choice for sequential assignment is triple resonances experiments such as HNCACB (Grzesiek and Bax, 1992b) and CBCA(CO)NH (Grzesiek and Bax, 1992a) which uses, as the name suggests, three different nuclei (^{1}H, ^{13}C, ^{15}N) that are correlated. The experiments are performed on double labelled (^{13}C, ^{15}N) protein samples.

1.4.4.1. HNCACB

The HNCACB experiment is a 3D experiment which correlates the amide ^{1}H, ^{15}N resonances with both the intraresidue Cα and Cβ resonances, and with the $^{13}C\alpha$ and $^{13}C\beta$ of the preceding residue as shown in Figure 1-18.

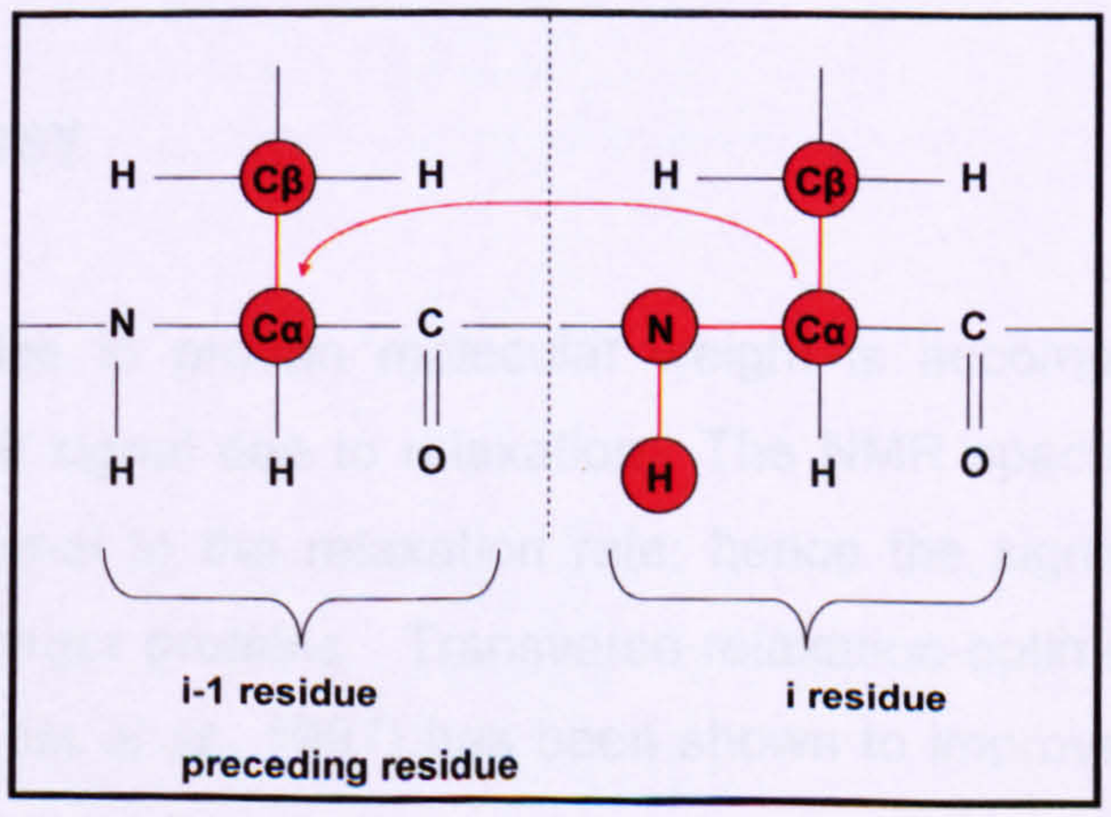

Figure 1-18. A dipeptide segment of a protein backbone showing correlations obtained in HNCACB experiments.

1.4.4.2. CBCA(CO)NH

The CBCA(CO)NH experiment is a 3D experiment which correlates the ^{1}H and ^{15}N amide resonances of one residue with both $^{13}C\alpha$ and $^{13}C\beta$ resonances of its preceding residue via the intervening ^{13}CO spin as shown in Figure 1-19. These experiments complement the HNCACB experiments and the two are used together for sequential assignment.

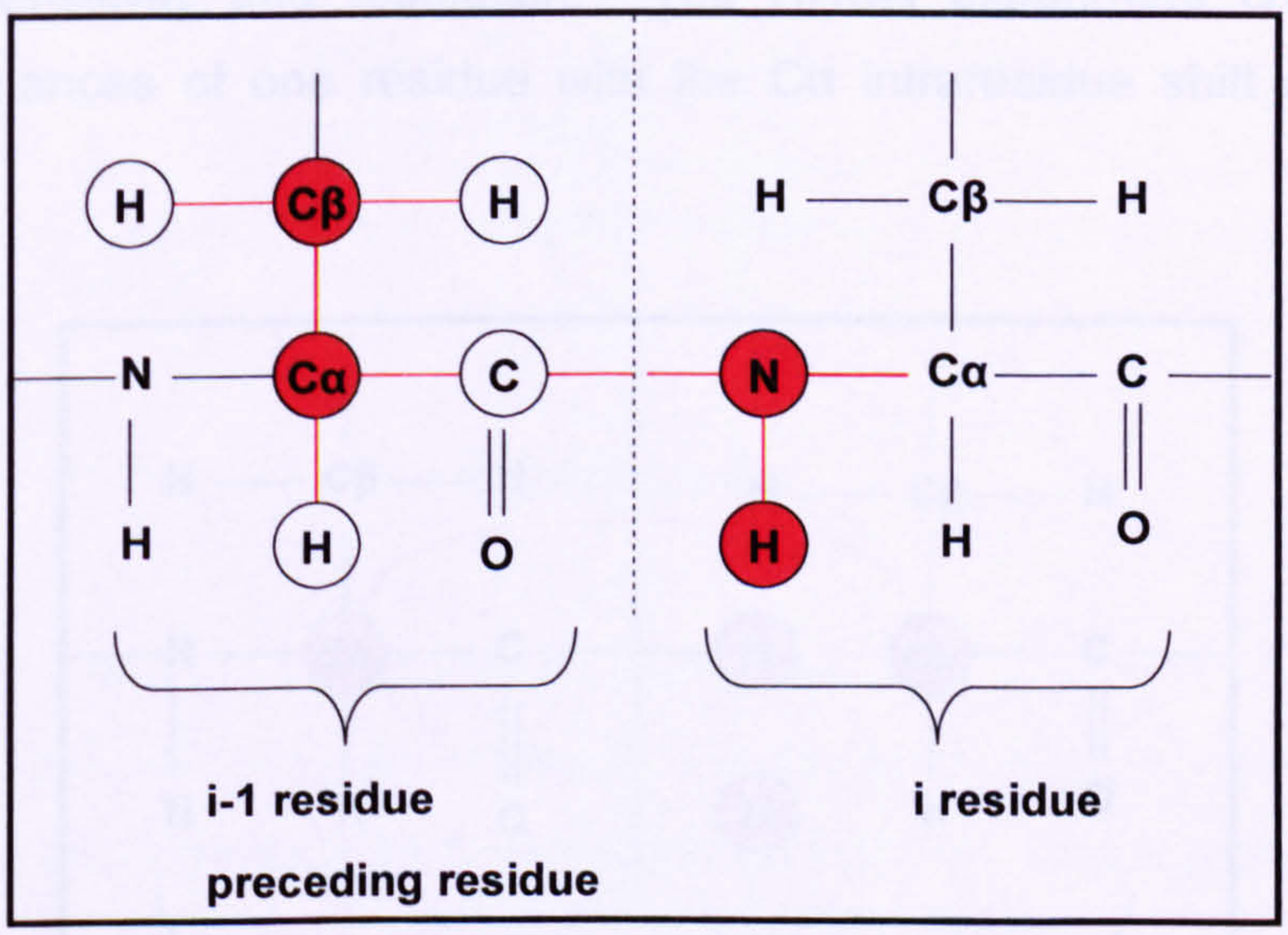

Figure 1-19. A dipeptide segment of a protein backbone showing correlations obtained in CBCA(CO)NH experiments.

1.4.4.3. TROSY

The increased size in protein molecular weight is accompanied with faster decay of the NMR signal due to relaxation. The NMR spectral line widths are inversely proportional to the relaxation rate; hence the signal-to-noise ratio is much higher for larger proteins. Transverse relaxation-optimised spectroscopy (TROSY) (Pervushin *et al.*, 1997) has been shown to improve the sensitivity of triple resonance experiments by suppressing the transverse nuclear spin

relaxation; the event that causes the deterioration of NMR spectra for larger molecular structures. The sensitivity can be increased with deuterated ^{13}C, ^{15}N labelled proteins (Salzmann *et al.*, 1998).

1.4.4.4. HNCA

The HNCA (Salzmann *et al.*, 1998) is also a 3D experiment, which in this research was carried out on a uniformly deuterated ^{13}C, ^{15}N labelled protein to maximise sensitivity and resolution. The HNCA experiment correlates ^{15}N amide resonances of one residue with the Cα intraresidue shift as shown in Figure 1-20.

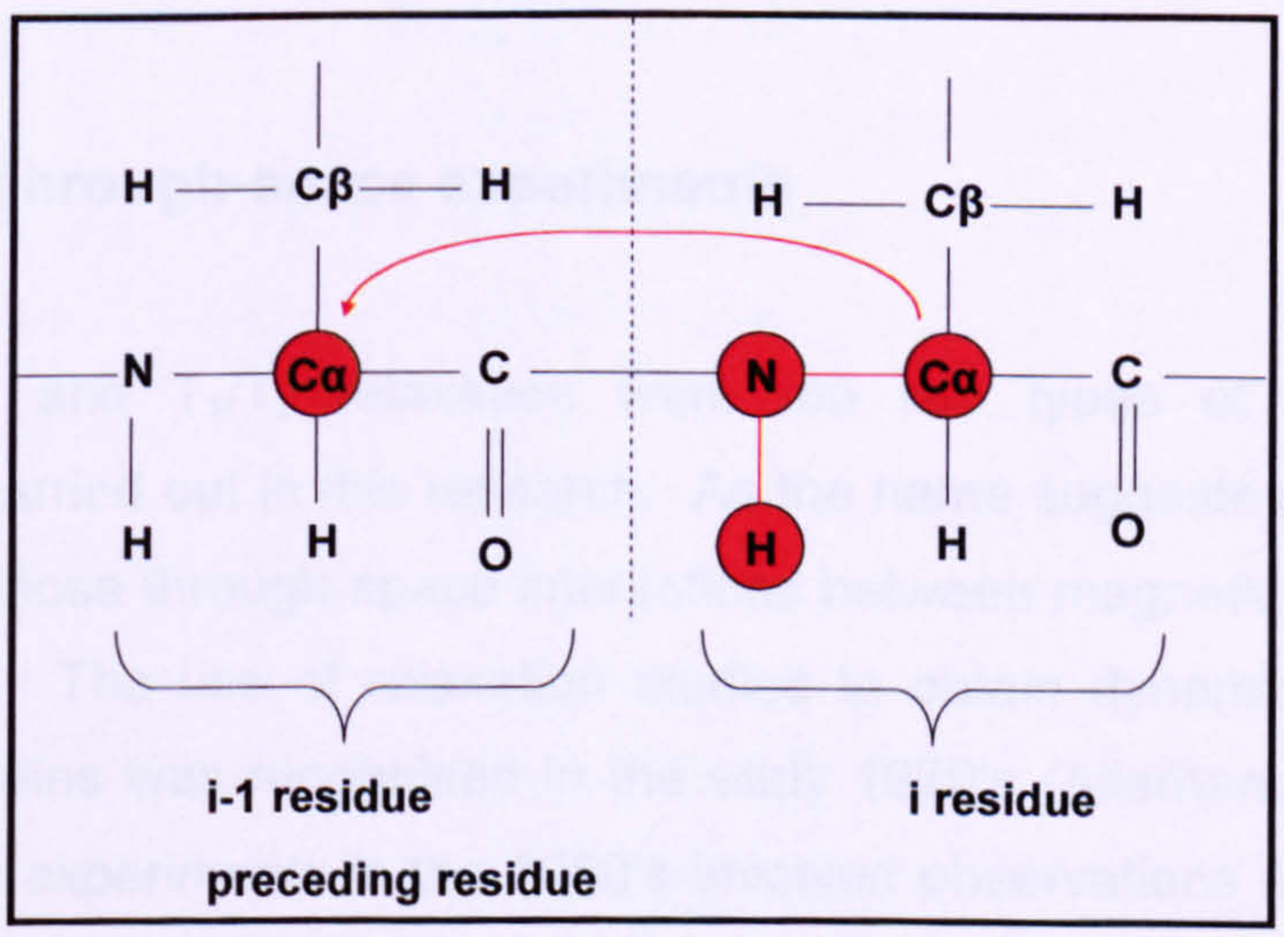

Figure 1-20. A dipeptide segment of a protein backbone showing correlations obtained in HNCA experiments.

1.4.4.5. TROSY - HN(CO)CA

The 3D HN(CO)CA (Salzmann *et al.*, 1998) is the complementing experiment for the HNCA experiment, used together for sequential assignment. This experiment correlates the ^{15}N amide resonances of one residue with the intraresidue Cα via the intervening ^{13}CO spin, shown in Figure 1-21.

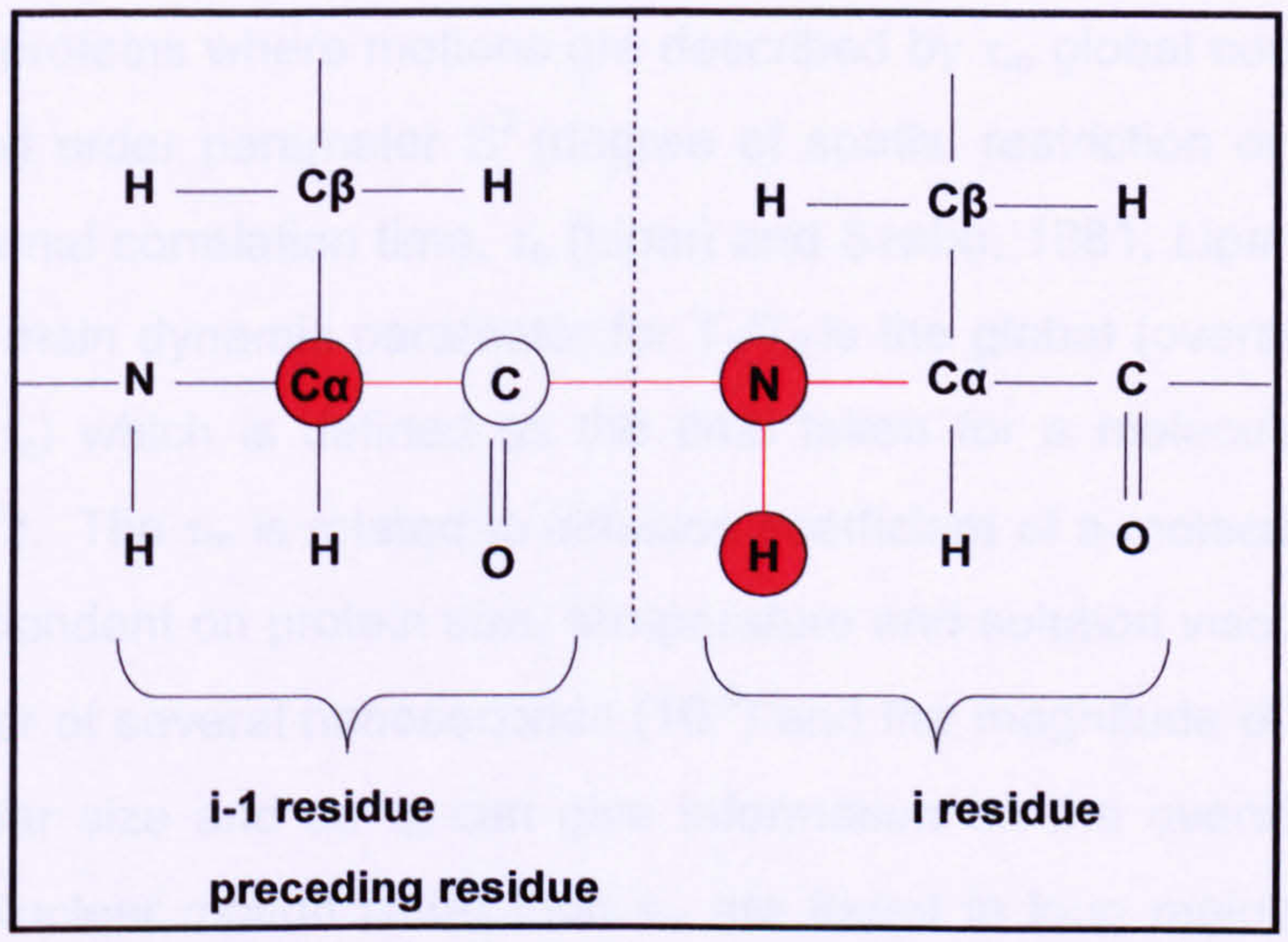

Figure 1-21. A dipeptide segment of a protein backbone showing correlations obtained in HN(CO)CA experiments.

1.4.5. Through-space experiments

The HetNOE and T_1/T_2 relaxation were the two types of through-space experiments carried out in this research. As the name suggests measurements recorded are those through-space interactions between magnetic dipoles of two nuclear spins. The use of relaxation studies to obtain dynamic properties of biological proteins was recognised in the early 1970's (Allerhand *et al.*, 1971). Early dynamic experiments in the 1980's involved observations made using 1D NMR techniques, due to lack of resolution of 1D experiments, dynamic also suffered from low resolution and sensitivity (Fuson and Prestegard, 1983; Smith *et al.*, 1987). More recently, now well established NMR relaxation experiments using 2D pulse schemes have been developed to provide higher resolution data pertaining to the internal dynamics of proteins (Kay *et al.*, 1989; Clore *et al.*, 1990). Data on the internal motional properties are important as they aid the understanding of protein functionality (Brunger *et al.*, 1987; Buchberger *et al.*, 2000).

Modern T_1/T_2 experiments have been analysed using the model-free formalism for isotropic proteins where motions are described by τ_m global correlation time, a generalised order parameter S^2 (degree of spatial restriction of motion) and effective internal correlation time, τ_e (Lipari and Szabo, 1981; Lipari and Szabo, 1982). The main dynamic parameter for T_1/T_2 is the global (overall) correlation time (τ_m or τ_c) which is defined as the time taken for a molecule to rotate 1 radian (57.3°). The τ_m is related to diffusion coefficient of a molecule in solution hence is dependant on protein size, temperature and solution viscosity. The τ_m is in the order of several nanoseconds (10^{-9}) and the magnitude of τ_m increases with molecular size and so τ_m can give information on the overall shape of a molecule. Nuclear motion faster than τ_m are found in loop regions and in the order of nano to picoseconds (10^{-9} to 10^{-12}). Side chain motions and movement of secondary structure elements are of slower order, milliseconds to nanoseconds (10^{-3} to 10^{-9}). These types of motions for single residues and globally for the whole molecule including size/overall shape can be analysed using the T_1 and T_2 relaxation measurements.

1.4.5.1. T_1 relaxation

Relaxation is the summation of the process which involves magnetisation decay over time. Heteronuclear longitudinal relaxations (T_1) also known as spin lattice relaxation involves redistributing the populations of the nuclear spin states in order to reach the thermal equilibrium distribution, the nuclear spin energy is lost to the surroundings or lattice. A π/2 pulse is applied which results in a magnetisation in the xy plane, the relaxation involves the return of the magnetisation with a time constant T_1 (see Figure 1-16). The rates of T_1 relaxation are strongly dependent on the NMR frequency and vary considerably with magnetic field strength, B_o. The T_1 rate is dependant on specific tumbling and internal protein properties, giving information on the size and shape of the protein; in terms of global τ_m on a nanosecond motion scale.

1.4.5.2. T_2 relaxation

Heteronuclear transverse relaxation time (T_2) also known as spin-spin relaxation, whereby nuclear spin coherence or energy is lost to a neighbouring spin system and is the decay constant for the component of M_z perpendicular (y axis) to M_o as shown in Figure 1-22. T_2 gives information on the nanosecond time scale, hence global motions but are affected by slower internal motions in the protein (τ_e). So the T_2 data give information on both the nanosecond time scale and slower internal motion on the micro-nanosecond time scale (10^{-6} to 10^{-9}).

The T_2 decay follows an exponential decay; T_2 is defined as the time required for the transverse magnetisation vector to return to 37% of the starting magnitude after the 90° excitation pulse. The two relaxation rates individually can be used to measure the backbone dynamics and together to analyse the anisotropy of molecular tumbling of the protein in solution.

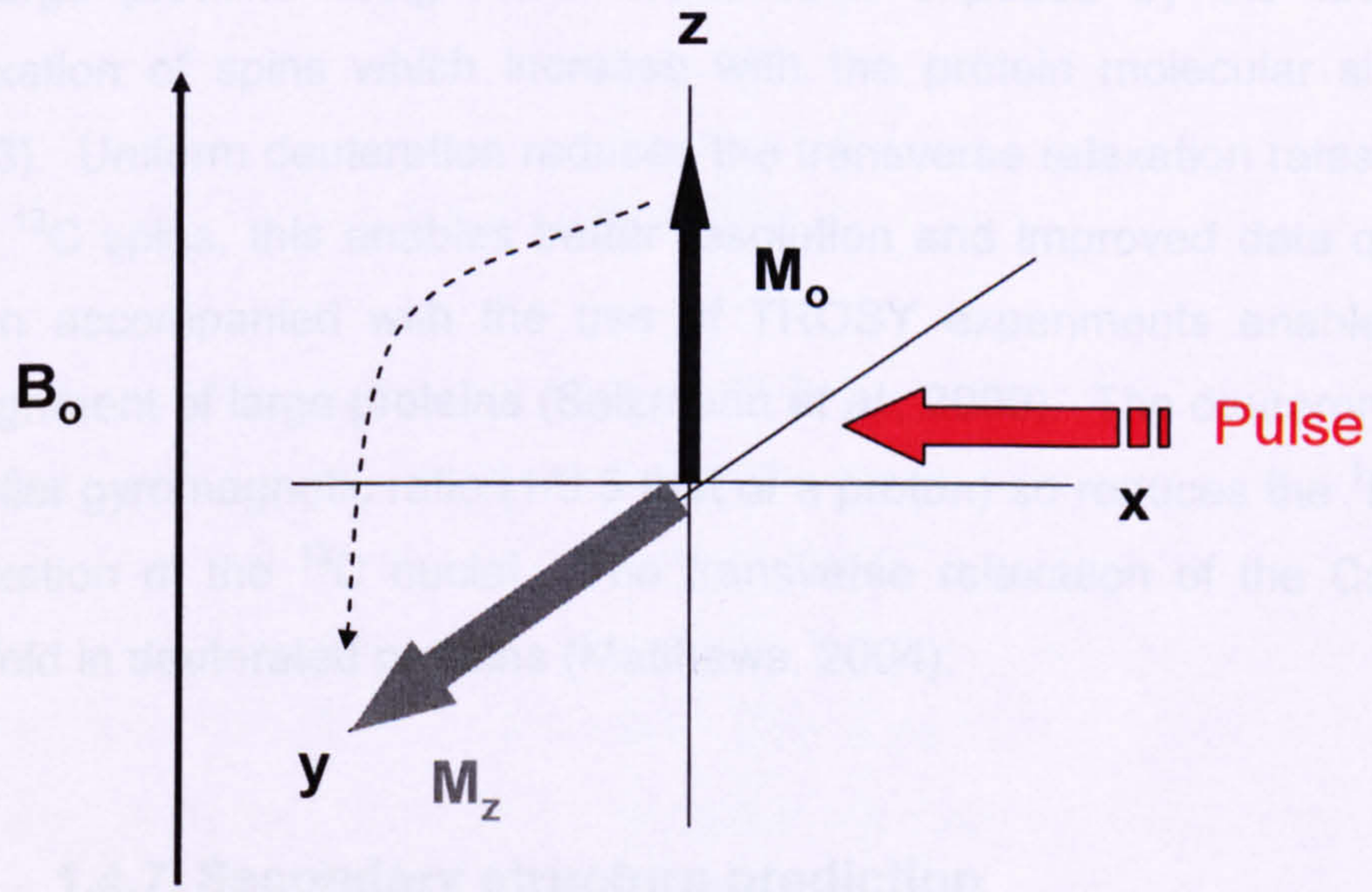

Figure 1-22. Diagram showing transverse magnetisation (T_2). z axis magentisation (M_o) in the same direction as the magnetic field (B_o), the transverse magentisation (M_z) is 90° on the y axis.

1.4.5.3. HetNOE

Heteronuclear Nuclear Overhauser Effect (HetNOE) is used to measure internal motions of proteins, providing information about the motion of individual N-H bond vectors. NOE's are observed between the amide proton and nitrogen NOE, and is the change in the intensity of an NMR resonance associated to cross-relaxation or transfer of magnetisation through space between atoms in close proximity and is more negative as fast picosecond motions increase. A decreased NOE intensity relative to the average observed for the majority of the residues indicates motion faster than the overall tumbling of the molecules. A standard HSQC is collected and then again with ^{1}H saturation to record ^{1}H - ^{15}N NOEs (Kay *et al.*, 1989).

1.4.6. Deuteration

In large proteins assignments have been impeded by the fast transverse relaxation of spins which increase with the protein molecular size (Wagner, 1993). Uniform deuteration reduces the transverse relaxation rates for ^{1}HN, ^{15}N and ^{13}C spins, this enables better resolution and improved data quality, which when accompanied with the use of TROSY experiments enabled backbone assignment of large proteins (Salzmann *et al.*, 2000). The deuteron has a much smaller gyromagnetic ratio (1/6.5 that of a proton) so reduces the ^{1}H-^{13}C dipolar relaxation of the ^{13}C nuclei. The transverse relaxation of the Cα is reduced fivefold in deuterated proteins (Matthews, 2004).

1.4.7. Secondary structure prediction

The Torsion Angle Likelihood Obtained from Shift and sequence similarity (TALOS) is a database system for empirical prediction of the phi and psi

backbone torsion angles (Cornilescu *et al.*, 1999). Torsion angles are important, since they are used in the Ramachandran plot which defines the helix and strand secondary structure and form the output of this program (Ramachandran *et al.*, 1963). The TALOS program uses a combination of five types of chemical shift (Cα, Cβ, $^{13}C'$, $^{1}H_N$ and Hα) for any number of residues in a given sequence. The program uses sequential groups of three amino acids of the input sequence which is interrogated against the database. This means for a central residue in a triplet, TALOS uses potentially 15 chemical shifts and three residue types. Hence if a triplet of residues in a structurally determined protein has similar chemical shifts to the target sequence, then the psi and phi angles of the structurally determined protein would be a very useful for predictions. The TALOS database holds 186 proteins with over 24000 triplets and when a target sequence is searched against this database the 10 best matches to the chemical shifts for every residue is outputted. Hence TALOS uses chemical shift data and sequence information to make quantitative predictions for protein backbone angles and to provide a measure of the uncertainties in these predictions.

1.4.8. Chemical shift mapping

Chemical shift mapping is a simple method of measuring changes in chemical shifts. These types of measurements are carried out using HSQC based experiments. Whereby a fully assigned HSQC is compared to another assigned HSQC and any perturbations recorded. The perturbations can then be mapped on to a structure or to a sequence. The perturbations resulting from chemical shifts are generally caused by changes in the residue local-environment or more specifically by temperature, ligand interactions or as in this research, interaction with adjacent domains. The inherit problem associated with chemical shift mapping is that any observed perturbations do not only result from direct

interaction but can also be observed in a site further away as a result of protein structural changes (Spitzfaden *et al.*, 1992).

1.4.9. Hydrogen/deuterium exchange

Hydrogen/deuterium (H/D) exchange experiments involves a process of dissolving a protein sample in D_2O, the backbone amide protons will, over time, exchange with deuterons. The intrinsic kinetics of the exchange process is dependant on pH, temperature and the neighbouring side chains (Woodward *et al.*, 1982; Englander and Kallenbach, 1983). The classical model for interpreting protein H/D exchange data was described by Linderstrom-Lang (Linderstrom-Lang, 1955). The model describes how the slow exchange hydrogens with the solvent occur as a result of structural fluctuations (local and global) around an average native conformation. This model therefore infers that hydrogen bonds are broken and reformed during these structural fluctuations and during these events proton transfer can take place. It was later experimentally proven that highly protected hydrogens had low exchange rates and are linked to thermal stability; only exchanging as result of global unfolding (Wuthrich *et al.*, 1980). Whilst intermediate exchanging hydrogen rates resulted from more localised structural fluctuations (Englander and Kallenbach, 1983). Hence H/D exchange experiments provide detailed information at an amino-acid specific level on protein structure, structure change and dynamics.

1.5. Aims of this Study

Since the discovery of PDI in 1964, it had taken 32 years before the first domain, the **a** domain, was structurally determined by NMR in 1996 (Kemmink *et al.*, 1996). The **b** and **a'** domains of human PDI were subsequently solved by NMR spectroscopy; x-ray crystallography attempts were made but had not been

fruitful. Protein flexibility or conformational exchange is widely accepted to cause difficulties when attempting to crystallise proteins. This is likely to be the cause of difficulties associated with PDI domains and x-ray crystallography attempts. Hence PDI domain structure determinations have been achieved using NMR where protein flexibility is not as great a hindrance to data collection. Therefore, a NMR based approach was deemed to be the best technique to use to derive structural data on the **b'** domain of human PDI.

It was not until two years into this research in 2006, that the ERp57 **bb'** domain combination was structurally determined by x-ray crystallography (Kozlov *et al.*, 2006), shortly after the NMR assignments of the thermophilic fungal *Humicola insolens* **b'** domain were published (Nakano *et al.*, 2006). Also that same year, and most importantly the first full length PDI structure from *Saccharomyces cerevisiae* (Yeast) was reported to have been solved by x-ray crystallography (Tian *et al.*, 2006), which was the first full length PDI related structure with functional activity. The only experimental data available on full length human PDI is derived from low resolution small angle x-ray scattering (SAXS) (Li *et al.*, 2006) also obtained in 2006.

So at the start of this research, the only data available was the structural data on the **a**, **b** and **a'** domains of human PDI, determined approximately ten years previous. The closest related protein to human PDI with a structure available was calsequestrin, but as highlighted earlier the **b'** domain shared very little sequence similarity to calsequestrin domains. Hence no related **b'** domain data was available and since the principal peptide binding site had been localised to the **b'** domain and therefore this domain was crucial for substrate binding; understanding this mechanism could be aided with structural determination of the final and only domain undetermined in human PDI. Furthermore, no structural data had, at this time, been reported on any PDI domain:substrate complexes which would provide valuable information on binding mechanisms.

The initial aim was structural determination of the **b'** domain, but during the course of the research it became apparent that there was clearly insufficient spectral resolution to achieve this goal, so only backbone assignment could be achieved. Prior NMR data suggested the inclusion of the **x**-region; this stabilises the **b'** domain and allowed the collection of better resolved spectra. The key outcome of obtaining full assignments of a domain construct including the **b'** domain would be to allow the study of dynamics and ligand binding properties. The ambitious aim of full structure determination was reduced to achieving the following aims:

- Assess and optimise the resolution of spectra collected on **b'** containing domain constructs.

- Obtain backbone assignments by collecting triple resonance multi-dimensional NMR experimental data, using isotopically labelled domain constructs which include the **x**-region.

- Obtain backbone dynamics data.
    - Heteronuclear relaxation T_1 and T_2 data.
    - Heteronuclear NOE data.
    - Hydrogen/Deuterium exchange.

- Use domain combinations to define domain interfaces.

Chapter 2. General Materials and Methods

2.1. Materials

All chemical reagents used were of analytical grade, and were obtained from the following sources unless otherwise stated, Fisher Scientific (UK), Sigma (USA). Deuterium oxide and isotopically labelled reagents were purchased from Cambridge Isotope Labs Inc. Celtone ^{2}H >97% and Spectra 9 $^{2}H/^{13}C/^{15}N$ >97% were purchased from Spectra Stable Isotopes (USA).

2.2. Protein Expression and Analytical methods

In this section the materials and methods used during the process of expressing and purifying various domain constructs of recombinant PDI proteins is discussed.

2.2.1. Expression constructs

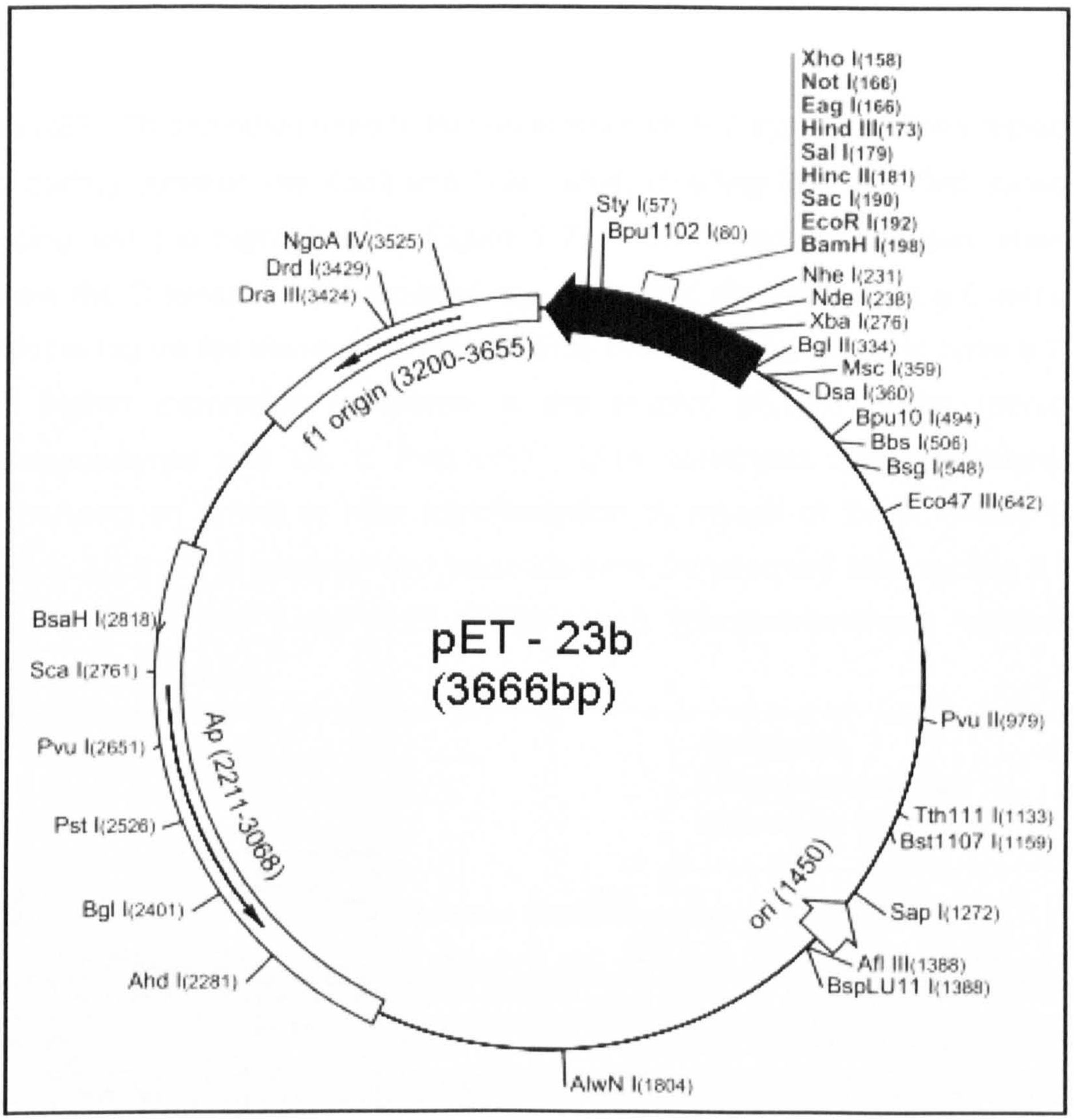

Figure 2-1. Vector map of pET-23b. This vector map was reproduced from the vector map for pET-23b from Novagen.

Expression vectors were derived from a pET-23b vector (see Figure 2-1). The expression vector encoding the various protein constructs, carrying a T7 RNA polymerase promoter and a multiple cloning site were kindly provided by the Dr. L. Ruddock group. Expression vectors were made by cloning into the Nde1 site, resulting in a protein with a non-cleavable hexa-histidine tag (MHHHHHHM). The non-cleavable N-terminal hexa-histidine tag was cloned into the Nde1 site, where the final HM forms the Nde1 site. The expression vectors which carry

ampicillin resistance were used for the expression of proteins with a hexa-his tag.

The pET-23b derivative used in this research included synthetic oligos replacing the coding between the Xba1 and Nde1 sites, resulting in a modified ribosome binding site (as highlighted in Figure 2-2). Also a stop codon was inserted before the C-terminal hexa-histidine tag to prevent the addition of a C-terminal histidine tag on the expressed protein. This construct was found to have a 2-10 fold higher expression compared to the original pET-23 vector (personal correspondence with Dr. L. Ruddock). DNA constructs were confirmed by sequencing on arrival or after transformation by myself or Dr. K. Wallis (see section 2.2.5, for procedure) and plasmids were transformed (see section 2.2.3, for procedure) into E.*coli* BL21 (DE3) pLysS (chloroamphenicol resistance) strain cells obtained.

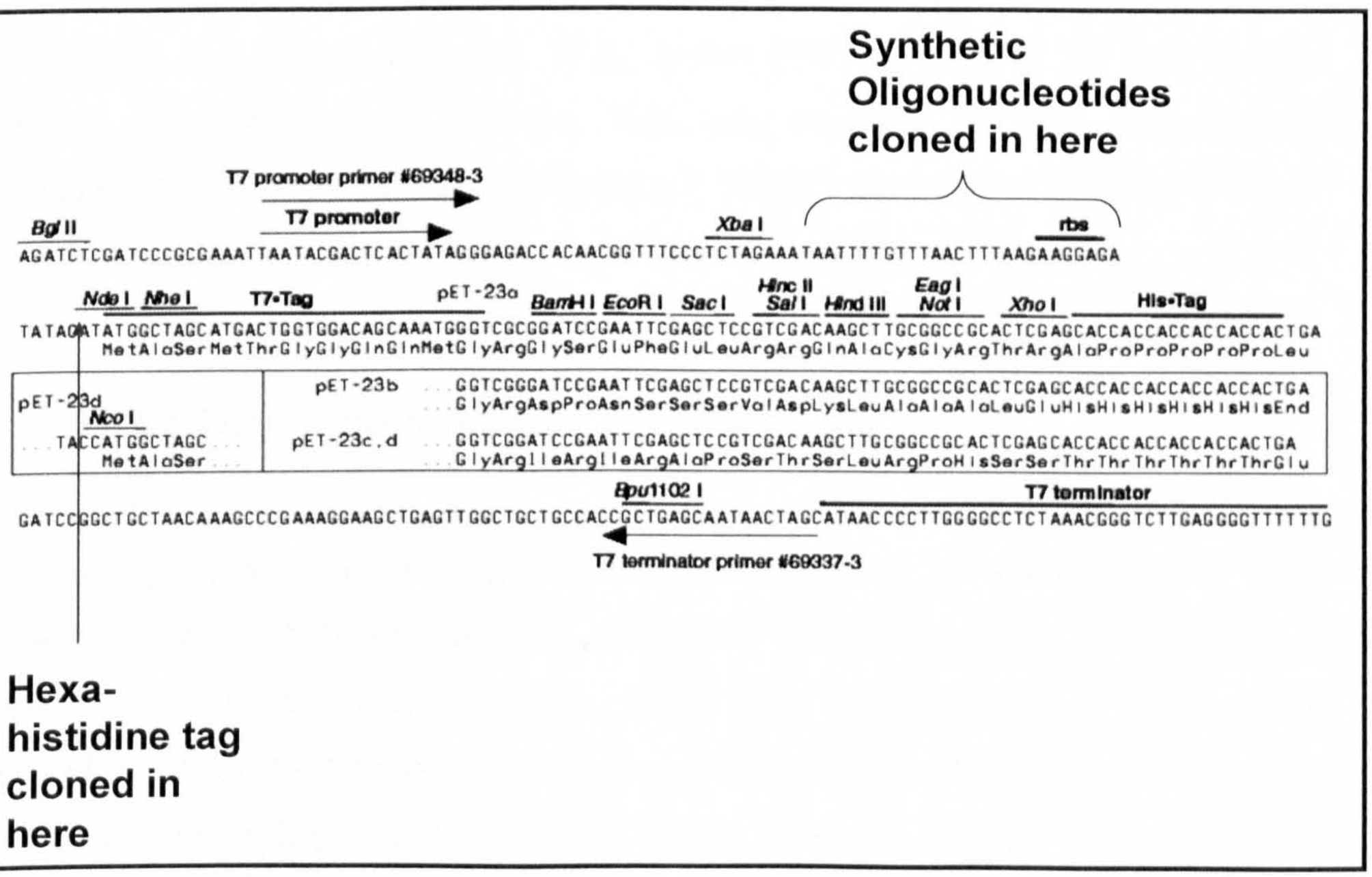

Figure 2-2. pET23b derivative cloning and expression region with highlighted regions where insertions were cloned in. Reproduced from the vector map for pET-23b from Novagen.

2.2.2. Preparation of E.*coli* competent cells for DNA transformation

LB agar plates with appropriate antibiotics added were streaked with the required E. *coli* strain (DH5α and BL21 (DE3) pLysS) from a glycerol stock (stored at -80°C) and incubated at 37 °C overnight. A single colony was used to inoculate a 10 mL sterile Luria-Bertani (LB) broth (with appropriate antibiotics added) and left shaking at 180 rpm overnight. 100 mL sterile LB broth was inoculated with 1.0 mL of overnight culture and incubated at 37 °C with shaking at 180 rpm until the optical density at 600nm reached 0.5. The culture was then centrifuged at 4000 x g for 10 minutes to pellet cells, and resuspended in TFB1 buffer (30 mM potassium acetate, 50 mM manganese chloride, 100 mM rubidium chloride, 10 mM calcium chloride, 15% (v/v) glycerol, pH 5.8), centrifuged and resuspended in TFB2 buffer (10 mM MOPS, 75 mM calcium chloride, 10 mM rubidium chloride, 15% (v/v) glycerol, pH 6.8), following the protocol adapted from Hanahan (Hanahan, 1985). Competent cells were frozen in 200 µL aliquots in liquid nitrogen and stored at -80 °C.

2.2.3. DNA transformation of E. *coli* cells

Chemically competent cells (prepared as described in section 2.2.2) were thawed on ice and 100 µL of cells incubated with approximately 50-200 ng of plasmid DNA for 30 minutes on ice. Cells were heat-shocked at 42 °C for 90 seconds and returned to ice for 2 minutes. 100 µL of sterile LB broth was added and cells were incubated at 37 °C with shaking for 60 minutes. The transformation mixture was then plated using aseptic technique onto LB agar plates containing the appropriate antibiotic selection for the transformed plasmid. Plates were incubated at 37 °C overnight.

2.2.4. Preparation of plasmid DNA

Plasmid DNA was extracted from E.*coli* cells grown overnight in LB containing the appropriate antibiotic, using a Qiagen miniprep kit, according to the manufacturer's instructions. These kits employ the alkaline lysis method, where DNA is purified by adsorption to a silica matrix under high salt conditions.

2.2.5. DNA sequencing of plasmid constructs

To ensure that the cloned gene had inserted correctly into the plasmid vector and no mutations were incorporated by PCR errors, constructs were sequenced using approximately 500 ng of plasmid DNA (prepared as described in 2.2.4) and 5 pmol of primer per sequencing reaction. Specific primers for the T7 promoter at the 5' end of the cloned gene (5' TAATACGACTCACTATAGG 3', Novagen) were used. Reactions were carried out using an ABI-Prism DNA sequencer, by the molecular biology service (Department of Biological Sciences, University of Warwick). The resulting sequences were translated into the protein sequence using the program ClustalW (Chenna *et al.*, 2003) to check for point mutations introduced by PCR. Sequences were also manually checked to identify the presence of start and stop codons and ensure the genes had inserted in the correct orientation. All PDI domain constructs were made with the domain boundaries as shown in Figure 1-5.

2.2.6. Preparation of E. *coli* glycerol stocks

5 mL sterile LB broth was inoculated with a single E. *coli* colony from a fresh transformation (described previously in section 2.2.3) and grown overnight in the presence of the appropriate antibiotic for plasmid selection with shaking at 37 °C.

600 µL of culture was aseptically mixed with 400 µL analytical reagent grade 50% (v/v) glycerol in a Corning cryovial and frozen at -80 °C.

2.2.7. Expression of recombinant protein in *E.coli*

Recombinant proteins were expressed in *E.coli* BL21 (DE3) pLysS strains. Cells which contain the expression plasmid were plated with the appropriate selective antibiotics (Chloroamphenicol and Ampicillin). A single colony was picked for a small scale 2 x 20 mL overnight culture at 37 °C shaking at 180 rpm. The small overnight culture was used to inoculate a larger culture, typically two 400 mL cultures, starting at $O.D._{600}$ 0.1; the volume of inoculant required was calculated as follows:

Volume of inoculant to start at expression at $O.D._{600}$ 0.1 (mL) =
Volume of expression culture (mL) / ($O.D._{600}$ Overnight Culture / start O.D.)

Cultures were then grown at 37 °C shaking at 180 rpm until $O.D._{600}$ reached 0.5, unless otherwise stated. Cultures were then induced with 1mM isopropyl β-D-thiogalactopyranoside (IPTG) and grown for 4 hours (varied depending upon optimum expression time) at 37 °C shaking at 180 rpm, unless otherwise stated. Cells were harvested by centrifugation at 15900 x g using a Beckman JA-8.100 rotor for 10 min at 4 °C. The cell pellet was then re-suspended in buffer A (20mM phosphate buffer, pH 7.3), in a tenth of the starting culture volume with final concentration of 10 mg/l DNase added and the sample was frozen. Once frozen, the sample was freeze thawed twice to lyse the cells. The cells where then centrifuged at 17418 x g for 20 min at 4 °C using a Beckman JA 25.50 rotor. The supernatant was then applied to an Immobilized Metal-Affinity Column (IMAC) column. Once purified, protein samples were stored frozen.

2.2.8. Determination of protein concentration

Protein concentration was determined using a colorimetric assay at 595 nm (BioRad). A 10 µL protein solution was added to 790 µL water and 200 µL BioRad reagent in a cuvette, mixed and the absorbance measured at 595 nm using an Amersham Biosciences Ultrospec 2000 UV-visible spectrophotometer. Samples were diluted and remeasured if the samples gave a reading outside the linear range of the assay (>0.6 at 595 nm). Protein concentration (mg/ml) was then calculated using the Beer-Lambert law. Alternatively protein concentration was measured by a direct measurement of UV absorbance at 280 nm using a Cary 100 spectrophotometer (Varian). Whereby the extinction coefficient (ϵ_{280}) was calculated by submitting the protein sequence to a web-based server called Protparam (available at http://www.expasy.ch/tools/protparam.html). The protein concentration was then calculated in mg/ml using the Beer-Lamberts law.

2.2.9. LB broth and agar

LB Broth per litre contained 10g tryptone, 5g yeast extract, 10g NaCl made up to volume with deionised water and autoclaved. When cool, appropriate antibiotics were added. LB agar was prepared as LB broth with the addition of 15g per litre agarose, autoclaved and when cool appropriate antibiotic added before pouring in to Petri dishes

2.2.10. Minimal growth medium

Minimal media was used to allow the introduction of selective nitrogen, carbon and deuterium isotopes into expressed proteins. The minimal media was prepared as described in Table 2.

Nutrient	Amount (mg)
$(NH_4)_2SO_4$ (a)	1000
Glucose (b)	4000
Na_2HPO_4	6800
KH_2HPO_4	3000
NaCl	500
Na_2SO_4	42
EDTA	50
$MnCl_2$	16
$FeCl_3$	5
$ZnCl_2$	0.5
$CuCl_2$	0.1
$CoCl_2$	0.1
H_3BO_3	0.1
$MgSO_4$	0.25
$CaCl_2$	44
d-Biotin	1
Thiamine	1

Table 2. Minimal medium nutrient amounts for a 1 litre of culture.
(a) For ^{15}N labelled samples the $(NH_4)_2SO_4$ was replaced with $(^{15}NH_4)_2SO_4$ (^{15}N 99% Cambridge Isotope Labs Inc.).
(b) For $^{15}N/^{13}C$ labelled samples the glucose was replaced with D-Glucose U-$^{13}C_6$ (^{13}C 99% Cambridge Isotope Labs Inc.) and $(NH_4)_2SO_4$ was replaced with $(^{15}NH_4)_2SO_4$ (^{15}N 99% Cambridge Isotope Labs Inc.).

In the instance when deuteration was required, H_2O was replaced with an D_2O (Cambridge Isotopes Labs Inc.) diluted with sterile H_2O to give the appropriate percentage of deuterated atoms. For high deuterium percentages, some of the minimal medium nutrients were required to be prepared in D_2O as opposed to sterile H_2O.

2.2.11. SDS-polyacrylamide gel electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was prepared using the Tris-Glycine buffer system (Laemmli, 1970) to identify

polypeptides under denaturing conditions. The discontinuous gel system consisted of a resolving gel at pH 8.8 and a stacking gel at pH 6.8. SDS-PAGE gels were produced with 12 or 16 % resolving gels depending upon the molecular weight of the protein to be resolved, overlaid with 5% (w/v) stacking gel. Gels were cast using a Mini Protean II gel kit from Bio-Rad. The composition of the resolving and stacking gel is given below in Table 3.

Gel type	Reagent	12 % Gel	16 % Gel
Separation Gel premix	Water	15.9	13.2
	Acrylamide (40%)	13.3	16.0
	1.5M Tris pH 8.8	10	10
	10% SDS	0.4	0.4
	APS (µL)	60	60
	Temed (µL)	6	6
Stacking Gel premix	Water	19.0	
	Acrylamide (40%)	3.7	
	1.5M Tris pH 6.8	3.75	
	10% SDS	0.3	
	APS (µL)	24	
	Temed (µL)	24	

Table 3. Composition of 12 and 16 % separation gel and stacking gel. All volumes in mL unless stated.

8 mL of the separation premix was used to make two gels and polymerised by adding 60 µL APS and 6 µL Temed. 4 mL of the stacking gel premix was used for two gels and polymerised using 24 µL of APS and 24 µL of Temed. 20 µL protein samples were prepared for analysis by mixing with 5 µL of 4 x reducing sample buffer (1M Tris-HCl pH 6.8, 40% glycerol, 0.8% SDS, 0.1% (w/v) bromophenol blue, 10 % β-mercaptoethanol) and heat denatured at 95 °C in a heating block for 5 minutes. 10uL of the samples (unless otherwise stated)

were then loaded into the stacking gel wells and electrophoresed using running buffer (25 mM Tris, 0.192 M glycine, 0.1% SDS pH 8.3). A low molecular weight marker (purchased from Amersham Biosciences) was also loaded in one lane, which contained Phosphorylase b (97.0 kDa), Albumin (66.0 kDa), Ovalbumin (45.0 kDa), Carbonic anhydrase (30.0 kDa), Trypsin inhibitor (20.1 kDa) and α-lactalbumin (14.4 kDa). Gels were run at 200 Volts for 40 minutes or until the gel front reached the base of the gel. Proteins resolved by SDS PAGE were stained for 1 hour in 50% methanol, 10% acetic acid, 40% water containing 10 mg L^{-1} Coomassie R-250 dye. Gels were then de-stained overnight at room temperature in the same solution in the absence of the Coomassie R-250. Gels were photographed and dried between two sheets of cellulose stretched over a frame.

2.2.12. PAGE under native conditions

PAGE run under native conditions involved the same method as SDS-PAGE under denaturing and reducing conditions (2.2.11), but omitting SDS and β-mercaptoethanol from the stacking and resolving gel; also from the loading and running buffer.

2.2.13. Reduction and carboxymethylation

Protein samples (typically of a 20μL volume) were firstly reduced with the addition of 5 μL of loading buffer containing 10% SDS, β-mercaptoethanol with a final concentration of 5% and then heated at 95°C for 5mins. Carboxymethylation was carried out by the addition of 7.5μL 1M iodoacetamide and incubated at room temperature in the dark for 20 minutes. Samples were then run on SDS-PAGE as described earlier.

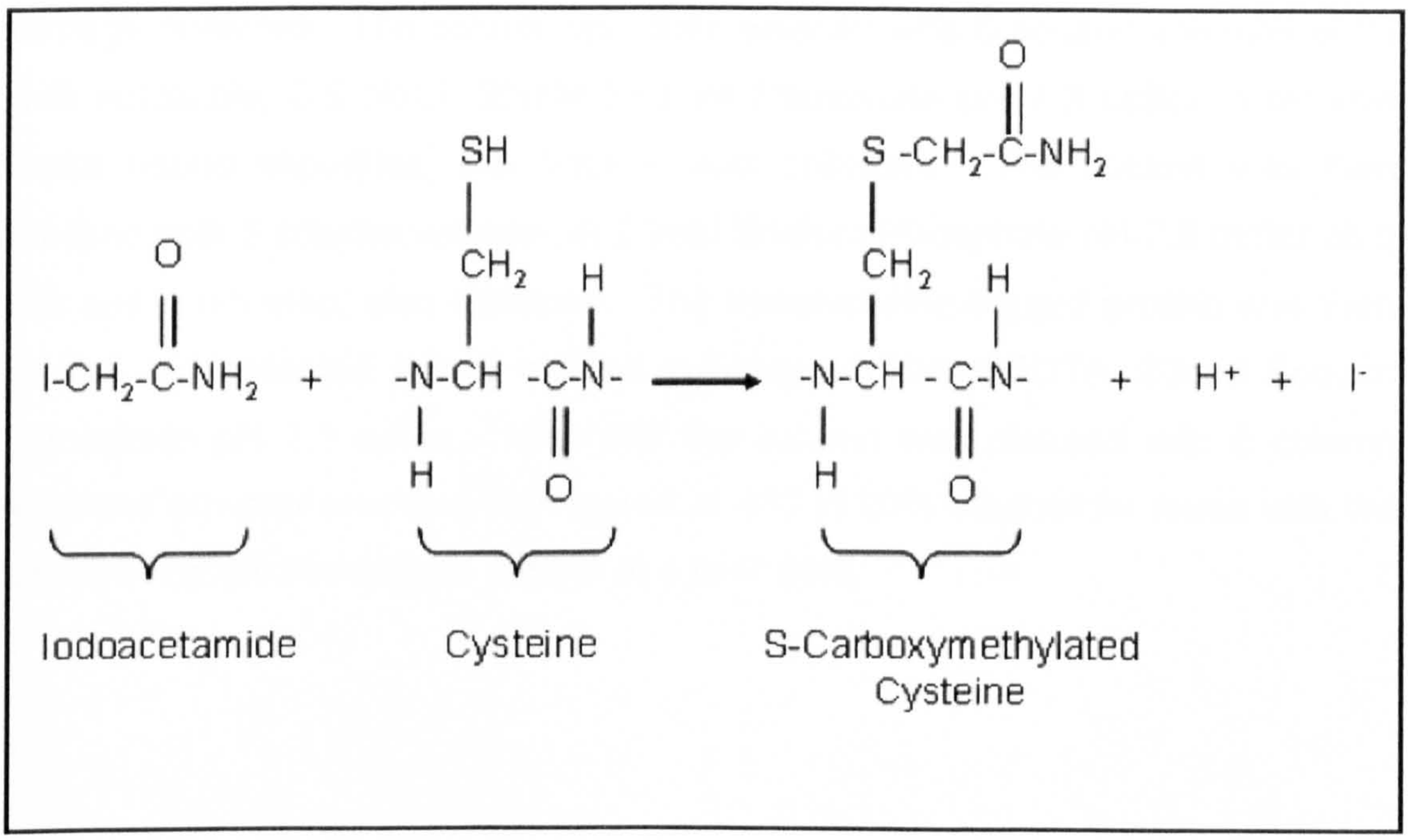

Figure 2-3. S-Carboxymethylation of cysteine residues scheme.

2.3. Chromatography Methods

2.3.1. Affinity chromatography

The first step of the purification of proteins was carried out using Immobilized Metal-Affinity Column (IMAC), where nickel chelated to chelating sepharose (Amersham Biosciences) resin was used to bind to the hexahistidine tags. Bench-top columns were produced by, packing glass wool into the bottom of an empty plastic 20mL syringe. Chelating sepharose was pipetted into the syringe to produce a column volume of 5mL. The resin was then washed with 10 column volumes of water and then 2mL of 0.2M nickel chloride added. The column was then washed with 5 column volumes of water and then 5 column volumes of 20 mM Sodium Acetate, 0.5M NaCl pH 3.0 to chelate the nickel to the sepharose resin. Excess nickel was then washed off and column re-equilibrated with 5 column volumes of 20mM Sodium Phosphate pH 7.3. The re-suspended clarified cell lysate was loaded onto the column and the flow-

through collected. The column was then washed with 5 column volumes of 25 mM Imidazole, 0.5 NaCl, 20mM Sodium Phosphate pH 7.3 buffer to remove loose bound impurities; this fraction was collected. The column was then washed with 5 column volumes of 20mM Sodium Phosphate pH 7.3 buffer as a low salt wash step; also collected. The hexahistidine-tagged protein was then eluted and collected with 5 column volumes of 50mM EDTA, 20mM Sodium Phosphate pH 7.3 buffer. After use the column was cleaned with 5 column volumes of water and was then stored at 4°C in 20% ethanol for reuse with the same hexahistidine-tagged protein at a later date.

2.3.2. Anion-exchange chromatography

Anion-exchange was carried out using a Source 30Q column (10 mL column with 30 µm beads) attached to the AKTApurifier 100 system (Amersham Biosciences). These resins have an attached quaternary amine group, which gives the resin a strong positive charge over the broad pH range of 2.0 to 12.0. Generally proteins are applied in a low salt buffer preferably above the protein isoelectric point, hence causing an overall negative charge on the protein and so maximise the binding to the positively charged resin. The protein can then be eluted with a salt gradient whereby weakly bound proteins are eluted in low salt and tightly bound proteins eluting at a high salt concentration.

Samples loaded onto the column, generally came from the affinity chromatography elution. The elute from the affinity chromatography was diluted from 25 mL to 100 mL in Buffer A (20 mM Sodium Phosphate pH 7.3), to reduce the salt concentration, before loading onto the anion-exchange column. Columns were equilibrated and washed in 5 column volumes of Buffer A and then 5 column volumes of 20 mM Sodium Phosphate, 0.5 M NaCl pH 7.3 (Buffer B) and finally back into 20 mM Sodium Phosphate pH 7.3 for a further 5 column volumes to ensure no excess salt was present. Samples were loaded via an injection loop and the column eluted by washing with a Buffer B gradient from zero to 100 % (0.5 M NaCl) over 10 column volumes, unless stated otherwise. Fractions of 2 mL were collected throughout the gradient and the presence of the eluted protein measured by an increase in absorbance at 280 nm.

2.3.3. Size exclusion chromatography

The Superdex 75 (Amersham Pharmacia Biotech, UK) gel filtration column matrix was poured as thick slurry into a 318 mL column (2.6 cm x 60 cm). The column was equilibrated in 2 column volumes of 20 mM sodium phosphate, 150

mM NaCl pH 7.3. Samples loaded on to the Gel filtration column were usually pooled fractions from the anion-exchange column and so pooled fractions were typically concentrated using Centricon filter columns (5000 Da molecular weight cut off) to give a final volume of 2 mL which could then be applied to a gel filtration column. The sample of interest was loaded via a 2 mL injection loop and the column was washed with the identical buffer with which the column was equilibrated at a flow rate of 2.0 mL per minute. The elution of excluded protein was monitored following the A_{280} on an AKTApurifier 100 system (Amersham Biosciences) and 3 mL fractions were collected between 0 mL – 300 mL.

2.4. Hydrodynamic methods

2.4.1. Analytical ultra-centrifugation – sedimentation velocity experiments

Sedimentation velocity experiments were performed using an XL-I analytical ultracentrifuge (Beckman Coulter) equipped with a four-cell An-60 Ti rotor at 20°C. A protein sample of 400 µL at a concentration of 79 µM in a 50mM Tris, 150mM NaCl buffer (pH 7.5) was used as the protein buffer and reference solution. The experiments were performed at a speed of 60,000 rpm for 3 hours. The sedimentation coefficients and molecular weight were calculated using SEDFIT (Schuck, 2000). These experiments were preformed by Lei Wang, in the lab of Prof. C.C. Wang (Chinese Academy of Sciences, Institute of Biophysics, Beijing).

2.5. Spectroscopic Methods

2.5.1. Electrospray mass spectrometry

For electrospray mass spectrometry undertaken by the Mass Spectrometry and Proteomics service (Department of Biological Sciences, University of Warwick), PDI protein samples were prepared to a concentration of between 10 – 200 µM in 50 µl 10mM ammonium bicarbonate. Samples were analysed by positive ion electrospray ionisation tandem mass spectrometry (ESI-MS/MS), to confirm the mass and relative isotope incorporation.

2.5.2. Circular dichroism spectroscopy

Circular dichroism spectra were recorded on a Jasco J600 spectropolarimeter, thermostatically controlled at 25°C. The protein samples were diluted to between concentrations 2.5 and 10 µM in 20mM phosphate pH 7.3 buffer and the final concentrations determined by spectrophotometric analyses at 280nm, using appropriate extinction coefficients and molecular weights. Far-UV spectra were recorded over a range of 190nm to 260nm in 0.5nm increments with a cell pathlength 0.1cm. Spectral data was averaged over 8 scans, corrected for baseline and the mean ellipicity per residue calculated. Spectra were corrected for buffer signal and converted to molar CD per residue before analysis using SELCON 3 algorithm (Sreerama *et al.*, 1999) on the DICHROWEB website (Lobley *et al.*, 2002) to estimate the secondary structure. Typical secondary structure signals are shown below in Figure 2-4.

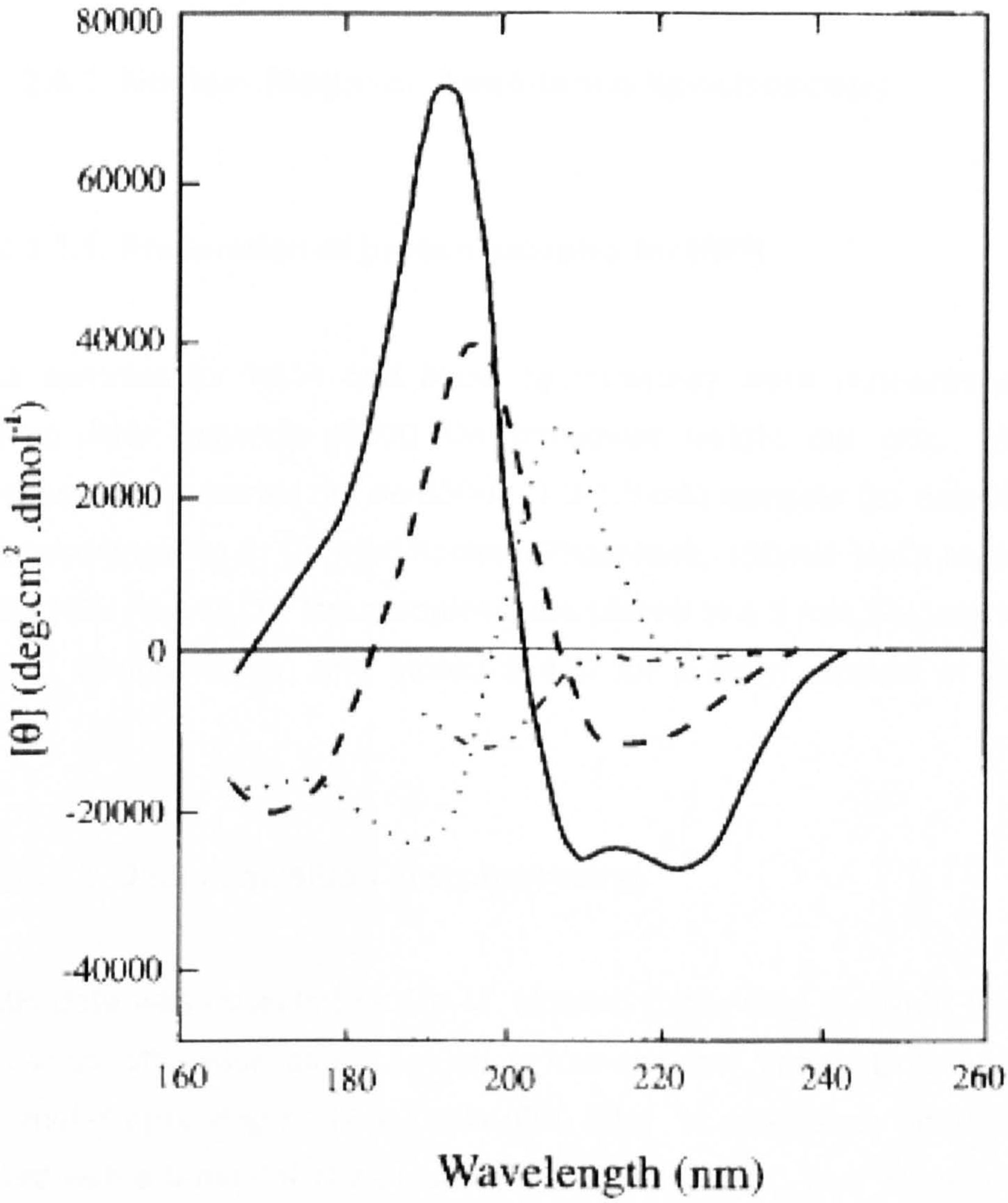

Figure 2-4. Typical secondary structure signals obtained by circular dichroism. Solid curve – α-helix, long dashes – anti-parallel β-sheet, dots – type I β-turn, dots and short dashes – irregular structure, taken from (Kelly and Price, 1997).

2.5.3. Nuclear Magnetic Resonance Spectroscopy

2.5.3.1. Preparation of protein samples for NMR

Protein samples for NMR and Mass Spectrometry were concentrated using Centricon filter columns (5000 Da molecular weight cut off). All NMR experiments were carried out on 330µL, 1.0-1.5 mM samples (as determined by A_{280} measurements) in 25 mM Sodium Phosphate, 150mM NaCl buffer at pH 6.5 with 10% (v/v) D_2O. The samples were placed in a 5 mm Shigemi matched with D_2O (Shigemi Inc.) and stored a 4°C for a short amount of time until required.

2.5.3.2. Data acquisition and processing

All NMR data was collected by Dr. M. Howard (University of Kent), typically at 25°C unless otherwise stated, using a four-channel Varian UnityInova NMR spectrometer operating at 14.1 Tesla (600 MHz ^1H resonance frequency) and equipped with a 5 mm HCN z-pulse field gradient probe. The ^1H chemical shift was referenced based on the position of the water resonance which has an exact value based on the relationship of $^1H_20$ resonances with temperature (Wishart and Sykes, 1994). The ^{13}C and ^{15}N were referenced using a spectrometer based macro which is the gyromagnetic ratios relationship (Wishart and Sykes, 1994). All experiments, unless stated, were solvent suppressed to reduce the water signal using WATERGATE (Piotto *et al.*, 1992) using a gradient field strength of 40-50 G cm^{-1}.

Data was processed using Varian and Bruker software, zero filled to increase the number of data points for the F_1, F_2 and when required in the F_3 dimensions.

In the triple resonance experiments a π/2.5 sine squared weighting function was applied in all dimensions prior to Fourier transformation. All NMR experiments and data processing were carried out by Dr. M. Howard (University of Canterbury, Kent). All data was analysed using the CcpNMR Analysis package (Vranken *et al.*, 2005), carried out by myself and with assistance from Dr. L. Byrne (University of Canterbury, Kent).

2.5.3.3. $^{15}N/^{1}H$ HSQC

The $^{15}N/^{1}H$ heteronuclear single quantum correlation (HSQC) experiments were collected with acquisition times were 45 ms in the F1 (^{15}N) and 114 ms in the F2 (^{1}H) dimensions. The total acquisition time was approximately 30 minutes. Data was typically acquired with 2048 points (9000 Hz) in the direct F2 dimension (^{1}H) and 4096 points (1850 Hz) in the F1 (^{15}N) dimension. The ^{15}N dimension was centred on 118ppm. If a $^{13}C/^{15}N$ sample was used then a modified $^{15}N/^{1}H$ HSQC was used which incorporates ^{13}C decoupling during both T_1 and T_2 acquisition periods.

2.5.3.4. Triple resonance experiments

Sequence specific backbone resonance assignments (N, NH, Cα and Cβ) were obtained for **bb'x** and **b'x**, by identifying the intra and inter residue connectivities in the HNCACB and CBCA(CO)NH spectra using methods described by Bax and Grzesiek (Grzesiek and Bax, 1992a; Grzesiek and Bax, 1992b). A $^{13}C/^{15}N$ sample of **bb'x** and **b'x** were used to collect the data.

The **bb'x** HNCACB and CBCA(CO)NH datasets were acquired with 1024 points (9000Hz) in the direct F3 dimension (^{1}H) and 60 points (10000Hz) in the direct F2 dimension (^{13}C) and 28 points (1850Hz) in the indirect F1 dimension (^{15}N). The number of transient and relaxation delays were set to give 63 hours, 7

minutes experimental time for CBCA(CO)NH and 61 hours and 48 minutes for the HNCACB. Carrier frequencies for these triple resonances experiments were set to 4.568 ppm, 45.743 ppm and 118.246 ppm for the ^{1}H, ^{13}C and ^{15}N respectively.

The **b'x** HNCACB and CBCA(CO)NH datasets were acquired with 768 points (9000Hz) in the direct F3 dimension (^{1}H) and 60 points (10000Hz) in the direct F2 dimension (^{13}C) and 32 points (1800Hz) in the indirect F1 dimension (^{15}N). The number of transient and relaxation delays were set to give 70 hours, 3 minutes experimental time for CBCA(CO)NH and 69 hours and 40 minutes for the HNCACB. Carrier frequencies for these triple resonances experiments were set to 4.568 ppm, 45.743 ppm and 118.246 ppm for the ^{1}H, ^{13}C and ^{15}N respectively.

Using the deuterated sample HNCA and HNCOCA experimental data was obtained where data was acquired with 2040 points (12000Hz) in the direct F3 dimension (^{1}H) and 48 points (2700Hz) in the direct F2 dimension (^{13}C) and 56 points (2818Hz) in the indirect F1 dimension (^{15}N). The numbers of transient and relaxation delays were set to give approximately 48 hours experimental time for HNCOCA and 48 hours for the HNCA. Carrier frequencies for these triple resonances experiments were set to 4.760 ppm, 117.558 ppm and 176.131 ppm for the ^{1}H, ^{13}C and ^{15}N respectively.

2.5.3.5. Sequential backbone assignment

Sequential assignment of the NMR backbone were achieved using HNCACB and CBCA(CO)NH experiments using a ^{13}C /^{15}N sample and HNCA and HNCOCA experiments using a deuterated sample. The approach is based on the unique Cα and Cβ chemical shifts of Glycine, Alanine, Serine and Threonine which differ from all other amino acid chemical shifts. This difference is

highlighted in Figure 2-5, which are the average chemical shift standards obtained for all amino acids (Wishart *et al.*, 1995a).

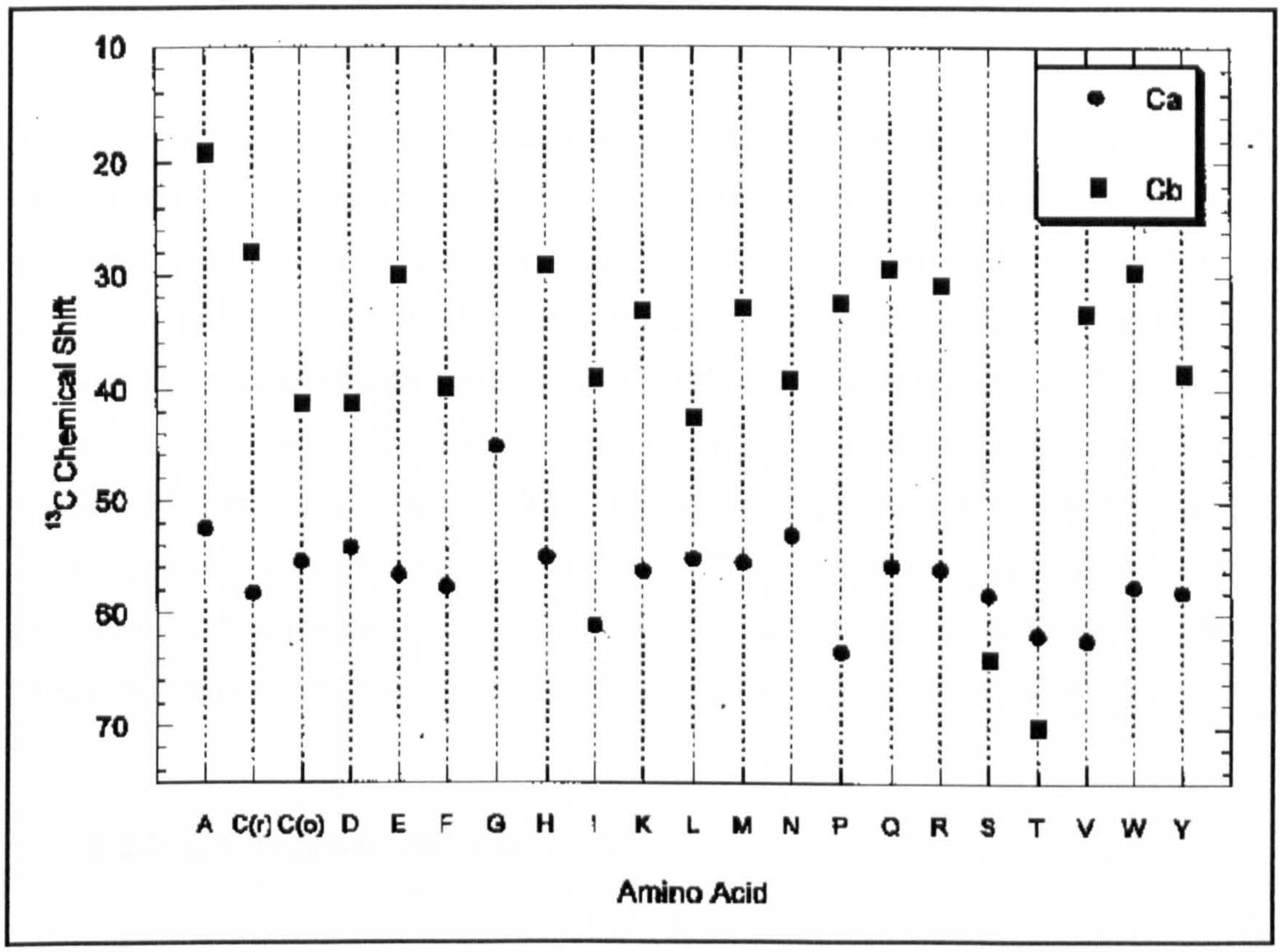

Figure 2-5. Chemical shift patterns for ^{13}Cα and ^{13}Cβ resonances for all amino acids in random coil conformation, adapted from (Wishart *et al.*, 1995a).

The method of sequential alignment involved matching the patterns of the Cα and Cβ of the i-1 amino acid in the CBCA(CO)NH data with the Cα and Cβ of the i amino acid in the HNCACB data. The Analysis software allowed each data pair, Cα and Cβ of a single amino acid to be interrogated against all other Cα and Cβ shifts, producing a list of matches in rank order of closest fit, which the analyst can use to link residues to produce a sequential assignment. Using the unique nature of certain residues mentioned previously, it is possible to unambiguously assign the ^{13}Cα, ^{13}Cβ, ^{15}N and ^{1}H resonances. In a similar fashion the HNCA and HNCOCA data was used to also sequentially assign residues, so when ambiguity or NMR overlap occurred, these assignments were

used to confirm/fill-in assignments. Also, other data sets were used to cross-correlated the assignment for **b'x** and the **b** domain.

2.5.3.6. Secondary structure prediction using TALOS

The chemical shift data for **bb'x** was used to obtain secondary structure predictions. The Analysis software has a facility to output the collated NMR chemical shift data in the appropriate format to be inputted into the TALOS program. TALOS is a UNIX based program which uses a tabulated Cα, Cβ, $^{13}C'$, $^{1}H_N$ and Hα chemical shift input file to predict secondary structure. The TALOS output gives a Ramachandran graphical interface where individual residue predictions can be analysed. The output highlights the secondary structure predictions in terms of helix, turn and coil. The prediction algorithm determines secondary structure by predicting the phi and psi angles of residues using the chemical shift data provided for that residue (Cornilescu *et al.*, 1999).

2.5.3.7. Chemical shift mapping

The chemical shift difference was calculated as shown below:

$$\text{Shift difference} = \sqrt{(\Delta^{1}H_{N})^{2} + {}^{1}/_{6}(\Delta^{15}N)^{2}}$$

The $\Delta^{1}H_N$ is the chemical shift change in the $^{1}H_N$ dimension (ppm) and $\Delta^{15}N$ is the chemical shift change in the ^{15}N dimension (ppm). The chemical shift changes are squared to ensure the difference is a positive number and ^{15}N difference is scaled down by 1/6 due to the chemical shift range of ^{15}N being approximately 30 ppm and the $^{1}H_N$ chemical shift range of approximately 5 ppm.

2.5.3.8. T_1 and T_2 relaxation experiments

All spectra was collected by Dr. M. Howard using a 1.5 mM ^{15}N-**bb'x** sample, in 25 mM Sodium Phosphate, 150mM NaCl buffer at pH 6.5 with 10% (v/v) D_2O at 25°C, using the four-channel Varian UnityInova NMR spectrometer operating at 14.1 Tesla (600 MHz ^{1}H resonance frequency) at the University of Kent NMR facility. To obtain T_1 relaxation rates delays of 128, 256, 385, 512, 641, 769 and 897 milliseconds were employed. To obtain T_2 relaxation rates delays of 20, 40, 60, 80, 100, 120, 140, 160 milliseconds were employed.

T_1 and T_2 values were extracted by measuring the peak volumes in the 2D-HSQCs as a function of a relaxation delay. Experimental T_1 values for each residue were fitted to the three parameter nonlinear fit to the following equation $I(\tau)=I_\infty -[I_\infty-I_o] \exp(-\tau/T_1)$, and T_2 values were likewise obtained from the three-parameters fits to the equation $I(\tau)=I_\infty +[I_\infty+I_o] \exp(-\tau/T_2)$ as described by Mackay and colleagues (MacKay *et al.*, 1996). In these equations, I_o and I_∞ are the cross-peak volumes at zero- and infinite-time, respectively. The non-linear curve fitting was carried out using the GraphPad Prism4 program (GraphPad Software, San Diego, CA).

2.5.3.9. Heteronuclear NOE experiments

Spectra was collected by Dr. M. Howard using a 1.5 mM ^{15}N-**bb'x** sample, in 25 mM Sodium Phosphate, 150mM NaCl buffer at pH 6.5 with 10% (v/v) D_2O at 40°C, 600 MHz Varian UnityInova NMR spectrometer at the University of Kent NMR facility. A pulse sequence was used which recorded the correlation spectra for measuring ^{1}H-^{15}H NOEs as described by Kay and colleagues (Kay *et al.*, 1989; Stone *et al.*, 1992). HSQC spectra were recorded in the presence and absence of the ^{1}H saturation. Experimental data was obtained where data was acquired with 2048 points (9000Hz) in the direct F2 dimension (^{1}H) and 192

points (3700Hz) in the direct F1 dimension (^{15}N). Carrier frequencies for these triple resonances experiments were set to 4.580 ppm and 117.257 ppm ^{1}H and ^{15}N respectively. The peak heights were recorded fro both spectra and NOE values obtained by comparing peak heights for $(I-I_o)/I_o$, where I is the peak height with ^{1}H saturation off and I_o ^{1}H saturation on.

2.5.3.10. Hydrogen/deuterium exchange experiments

Spectra were collected by Dr. M. Howard using a 1.5 mM ^{15}N-**bb'x** and a ^{15}N-**b** domain sample in 25 mM Sodium Phosphate, 150mM NaCl buffer at pH 6.5, which were freeze-dried and dissolved in 100% D_2O. The first time point of 5 minutes was the time taken to dissolve the sample and collect the first HSQC at 25°C using the 600 MHz Varian UnityInova NMR spectrometer at the University of Kent NMR facility. H/D exchange values were extracted by measuring the peak volumes in the 2D-HSQCs as a function of time. The values were fitted to a single exponential decay curve and the decay rate calculated using the program GraphPad Prism4 (GraphPad Software, San Diego, CA).

2.6. Structure Modelling

2.6.1. Molecular modelling

Geno3D (http://geno3d-pbil.ibcp.fr), an automatic web server for protein molecular modelling was used to model the structure of the **b'x** domain. The protein sequence is supplied to the server which performs the homology modelling, firstly the program identifies homologous proteins with known 3D structures by using PSI-BLAST; in the case of **b'x** the structure from ERp57 **bb'** domain was used (PDB code 2H8l). The program then performs an alignment of both query and subject sequences, extracts geometrical restraints (dihedral angles and distances) for corresponding atoms between the query and the template. Then finally performs the 3D construction of the protein by using a distance geometry approach which was sent via e-mail to the user (Combet *et al.*, 2002).

Chapter 3. Preparation and Preliminary Characterisation of PDI Proteins

3.1. Introduction

Preliminary data generated by previous work in the group had indicated that the **bb'x** construct gave useful NMR spectroscopic data. The main objective of this thesis was therefore to undertake a detailed structural characterisation of the **bb'x** domain combination using NMR techniques; the preparation, optimisation and characterisation of labelled and unlabelled samples are described in this Chapter. As an input to this analysis, it was necessary to generate and analyse labelled samples of the **b** domain which previously had been characterised (Kemmink *et al.*, 1999). During the process of structural characterisation it was identified that **b'x** assignments could assist the **bb'x** assignments, this Chapter describes the preparation and characterisation of labelled **b'x** samples for NMR analysis.

In order to understand the the properties of the **bb'x** domains of PDI, it was necessary to compare these properties with those of full-length PDI and those of isolated **b** and **bb'** domains. In this Chapter, the preparation and preliminary characterisation of all these domains and domain combinations are reported. All domains and domain combinations were expressed in *E.coli* BL21 (DE3) pLysS cells prepared as described in the Material and Methods Chapter. Expression was carried out as described in section 2.2.7, unless otherwise stated.

Purification generally involved three steps, affinity chromatography, anion-exchange chromatography and finally gel-filtration; any deviations from this protocol are stated. Two species were identified during purification, it is

important to note that fractionation by anion-exchange chromatography led to two pools (Pool 1 and Pool 2). The early or first peak from anion-exchange chromatography was designated Pool 1 and the late or second peak was designated Pool 2. When these pools are further purified by gel-filtration, fractions are now referred to as dimer (early peak) and monomer (late peak).

3.2. b domain

The **b** domain had previously been characterised (Kemmink *et al.*, 1999). Early in to this study it was identified that the assignment of **bb'x**, expected to be demanding, could be aided with a preassigned spectrum of the **b** domain. Hence producing a ^{15}N-labelled **b** domain was essential to assess the quality of NMR data achievable and lead onto the backbone assignment of the **b** domain.

3.2.1. Expression and characterisation of the b domain

The **b** domain construct was grown as described in section 2.2.7. Two 400mL unlabelled minimal cultures were grown in two 2 Litre flasks (preparation described in section 2.2.10). The sample was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1.

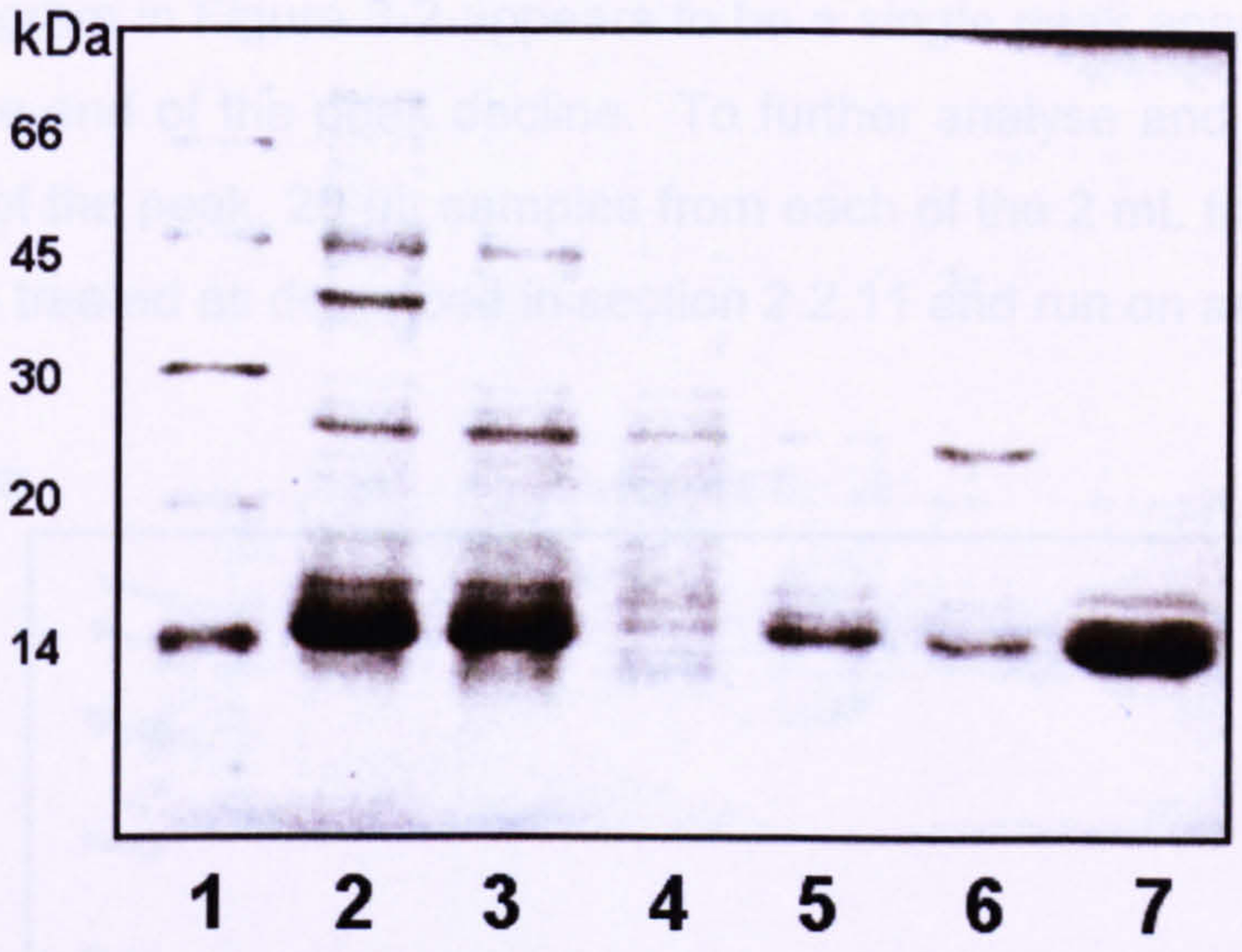

Figure 3-1. 12% SDS-PAGE of the b domain purification using IMAC. Lane 1, low molecular weight marker. Lane 2, Total expressed protein. Lane 3, Soluble protein fraction. Lane 4, IMAC Flow through. Lane 5, IMAC Wash fraction. Lane 6, Low salt wash fraction. Lane 7, hexa-histidine tagged b domain elute fraction.

It is clear that IMAC purification step in Figure 3-1 was very effective at removing virtually all of the non-tagged protein impurities, except one band of a slightly larger molecular weight, just above the main protein band in Lane 7. To ensure all impurities were removed, the elute was then prepared and further purified on an anion-exchange column as described in section 2.3.2

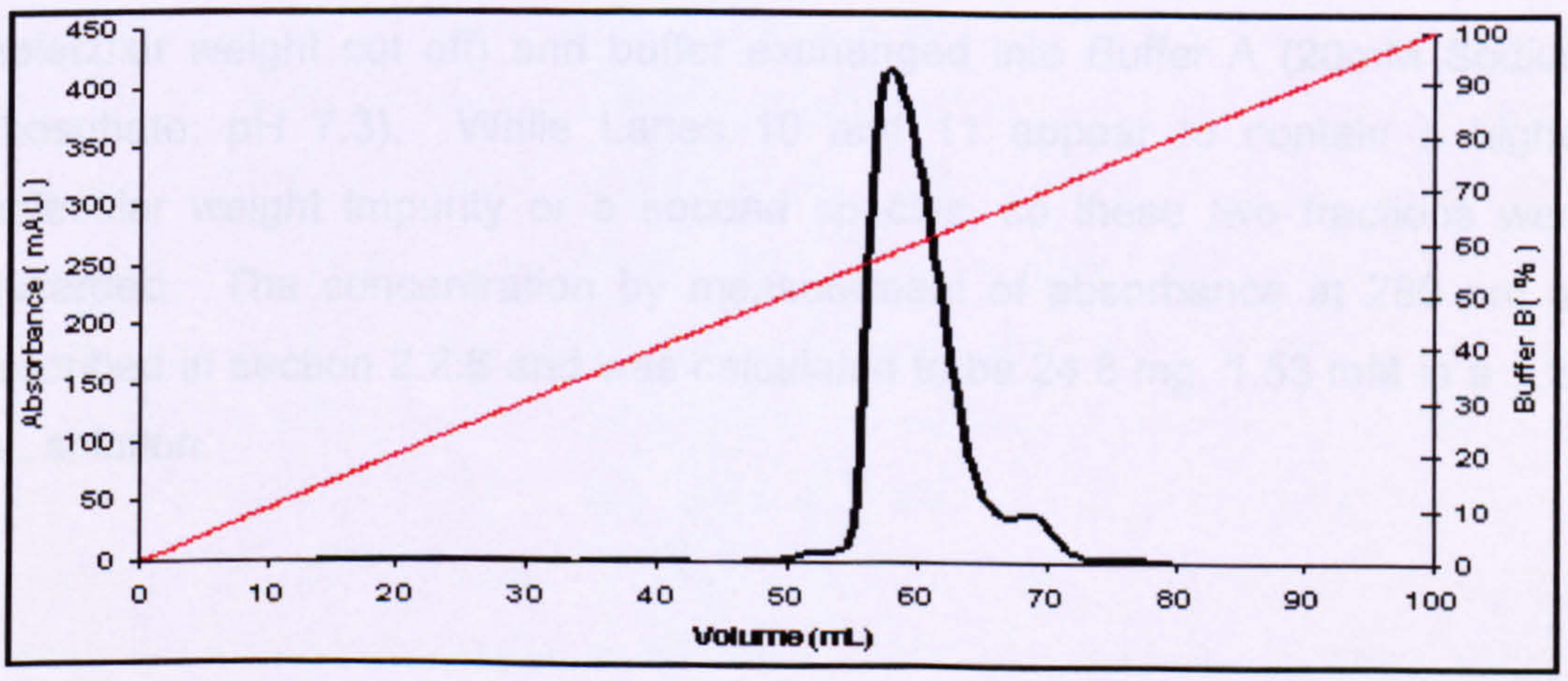

Figure 3-2 Chromatogram of b domain using a Source 30Q column. Black line is the absorbance at 280 nm, red line is the Buffer B 0 % to 100 %. The x axis shows the accumulative volume.

The chromatogram in Figure 3-2 appears to be a single peak apart from a small shoulder at the end of the peak decline. To further analyse and determine the homogeneity of the peak, 20 µL samples from each of the 2 mL fractions across the peak were treated as described in section 2.2.11 and run on an SDS-PAGE.

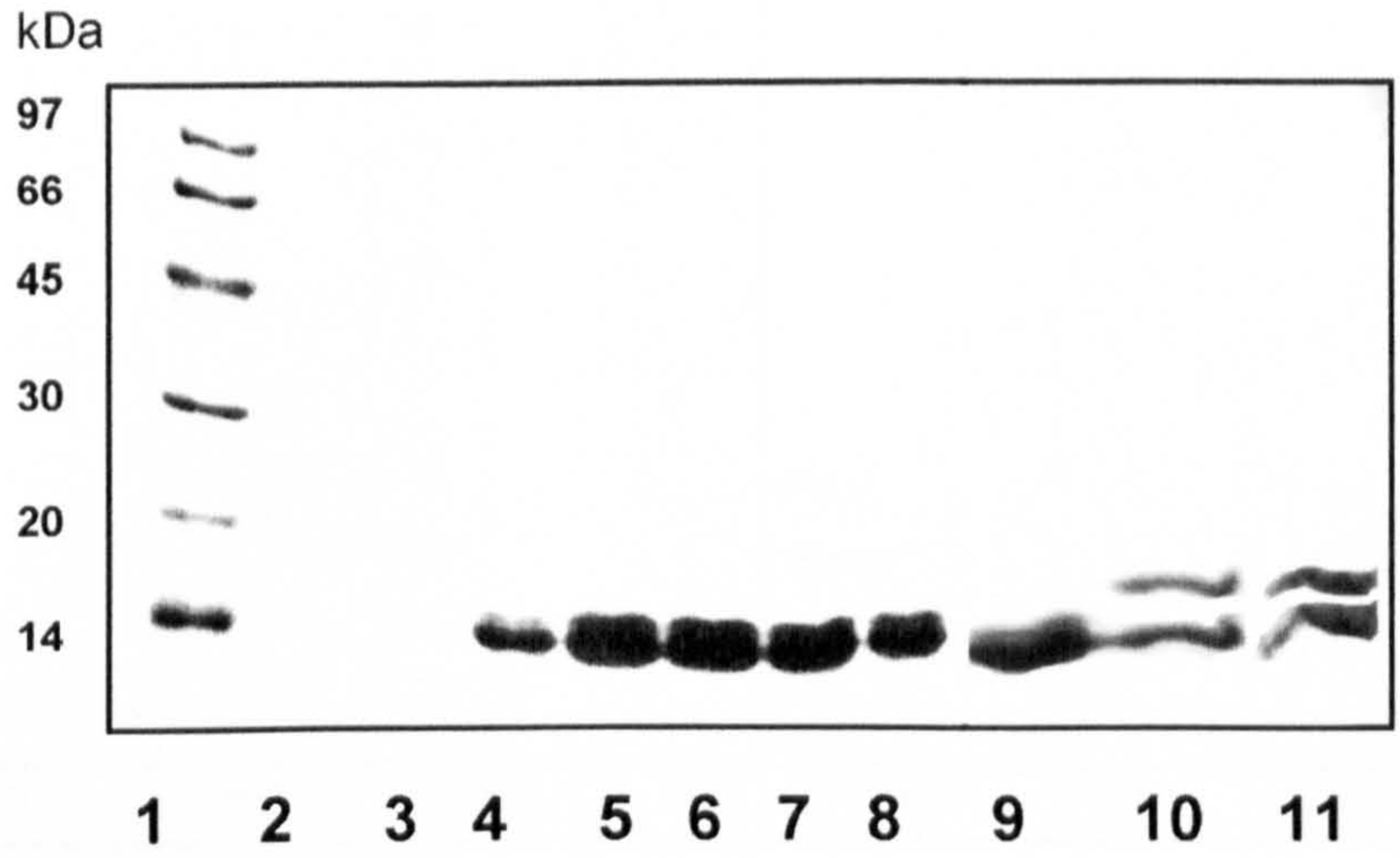

Figure 3-3. 12% SDS-PAGE of anion-exchange fractions of b domain. Lane 1, low molecular weight marker. Lane 2, fraction from 50-52 mL. Lane 3, fraction from 52-54 mL. Lane 4, fraction from 54-56 mL. Lane 5, fraction from 56-58 mL. Lane 6, fraction from 58-60 mL. Lane 7, fraction from 60-62 mL. Lane 8, fraction from 62-64 mL. Lanes 9, fraction from 64-66 mL. Lane 10, fraction from 66-68 mL. Lane 11, fraction from 68-70 mL.

The material in Lanes 4 to 9 in Figure 3-3 appear to be homogenous and so were pooled, concentrated down using a Centricon filter columns (5 kDa molecular weight cut off) and buffer exchanged into Buffer A (20mM Sodium Phosphate, pH 7.3). While Lanes 10 and 11 appear to contain a higher molecular weight impurity or a second species, so these two fractions were discarded. The concentration by measurement of absorbance at 280 nm as described in section 2.2.8 and was calculated to be 24.8 mg, 1.53 mM in a 1.25 mL solution.

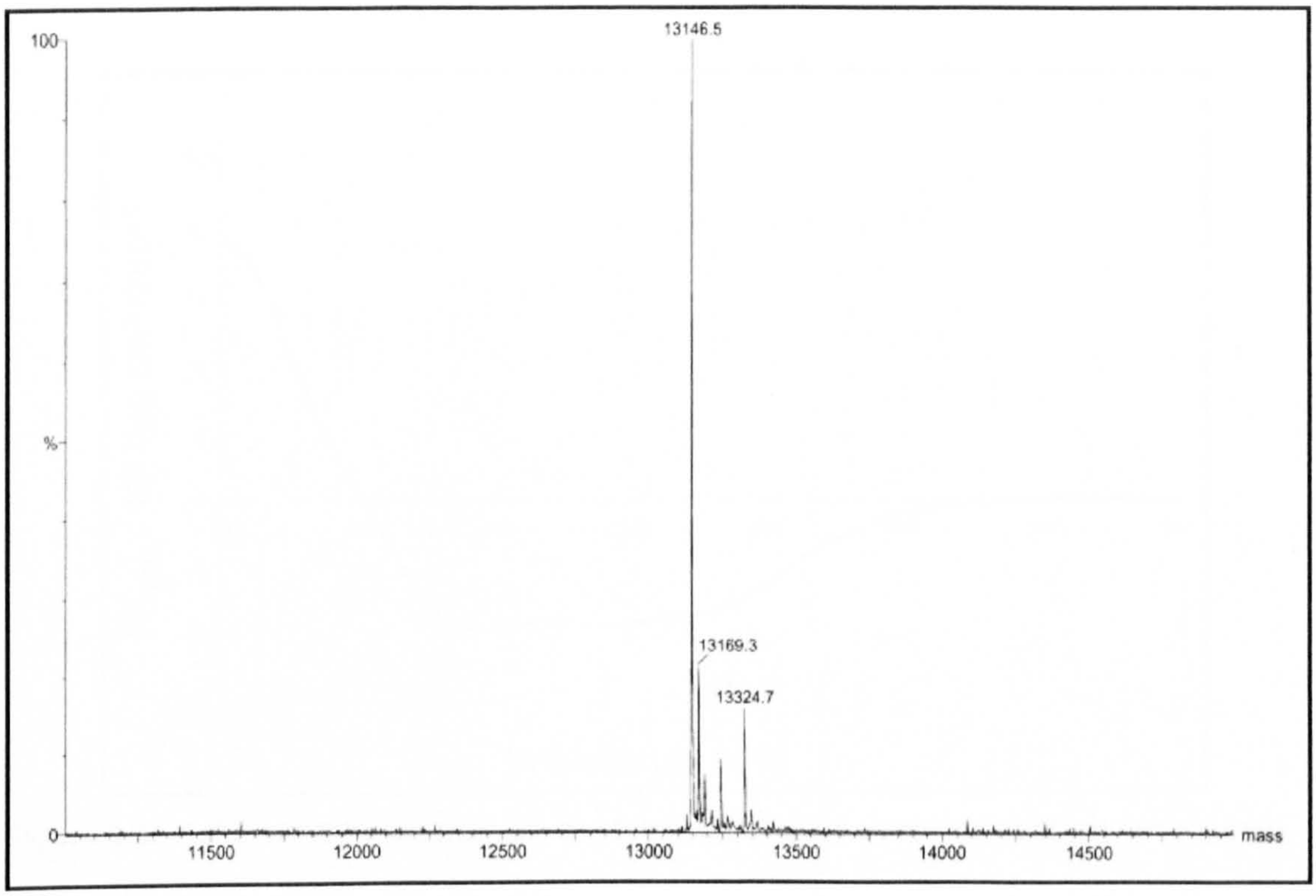

Figure 3-4. Positive ion mass spectrum of the b domain.

The deconvoluted electrospray mass spectrum gave a mass of 13146.5 Da shown in Figure 3-4, which was in agreement with the predicted mass using the sequence of 13147 Daltons. The species at 13169.3 Da, with an extra mass of 23 Da results from a sodium adduct. The other ion present was a species with an extra mass of 178 Daltons with a total mass of 13324.7 Da. This species results from a spontaneous α-N-6-Phosphogluconoylation of the histidine tag which is reported to produce a species with an extra mass of 258 or 178 Daltons. It has been show in *E.coli* that an extra mass of 258 Daltons results from a modification of the poly-histidine tag sequence involving the addition of a 6-phosphogluconoyl group and if this group is dephosphorylated, this results in an extra mass of 178 Daltons. A possible mechanism is the acylation by 6-phosphoglucono-1,5-lactone, produced from glucose-6-phosphate dehydrogenase (Geoghegan *et al.*, 1999). It was therefore possible that the slight double banding appearing in Lanes 4-9, in Figure 3-3 results from this histidine tag modified species.

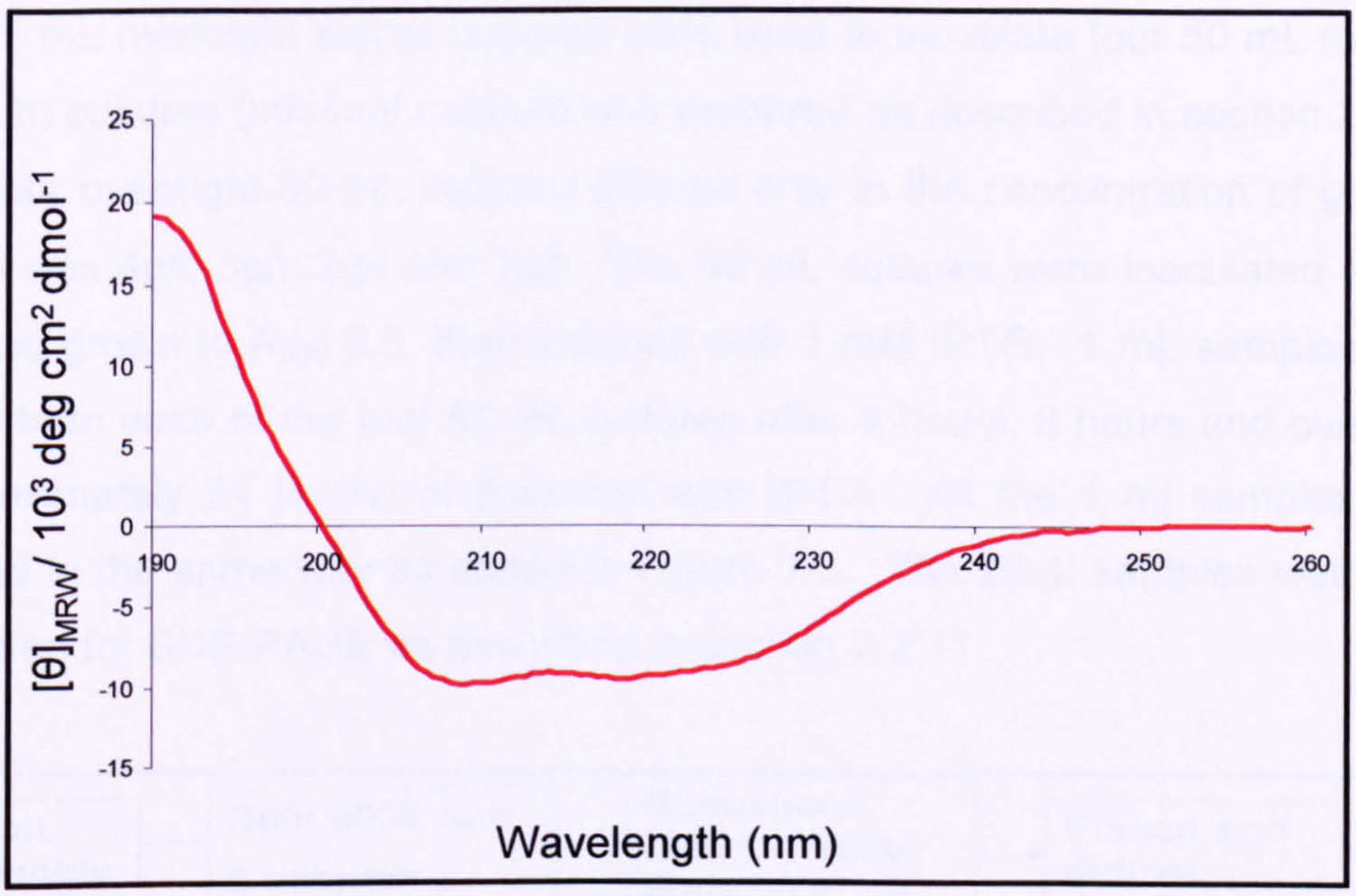

Figure 3-5. A Far UV CD spectrum for the b domain, using a sample of 10μM, 0.14 mg/ml.

From the CD spectrum shown in Figure 3-5 it was clear that the recombinantly expressed **b** domain was folded, since there was a large positive peak at approximately 190nm and two negative peaks at approximately 209 and 222 nm; typical of proteins with well-ordered structures. Analysis of the spectrum using the DICHROWEB website (Lobley *et al.*, 2002), indicated 30% helical and 18% beta content, which is in close agreement to the 34% helical and 18% beta content as quoted in the **b** domain (1BJX) PDB file.

3.2.2. Optimisation of b domain expression

The aim of this optimisation experiment was to identify the optimum induction time and attempt to reduce the amount of glucose used and yet still produce sufficient amounts of recombinant protein. A future aim is to produce ^{13}C-labelled protein for NMR experiments. Isotopically labelled ^{13}C-glucose was an expensive reagent and so here we aim to produce sufficient quantities of protein on the least amount of glucose supplied to the bacterial expression culture.

Four 5 mL overnight starter cultures were used to inoculate four 50 mL minimal medium cultures (minimal medium was prepared as described in section 2.2.10). The four overnight-50 mL cultures differed only in the concentration of glucose which was 4g/l, 3g/l, 2g/l and 1g/l. The 50 mL cultures were inoculated at A_{600} 0.1 and grown to A_{600} 0.5, then induced with 1 mM IPTG. 1 mL samples were taken from each of the four 50 mL cultures after 4 hours, 6 hours and overnight (approximately 24 hours) of induction with IPTG. All the 1 ml samples were treated in the same way as stated in Figure 3-6. The 20uL samples were then prepared for SDS-PAGE as described in section 2.2.11.

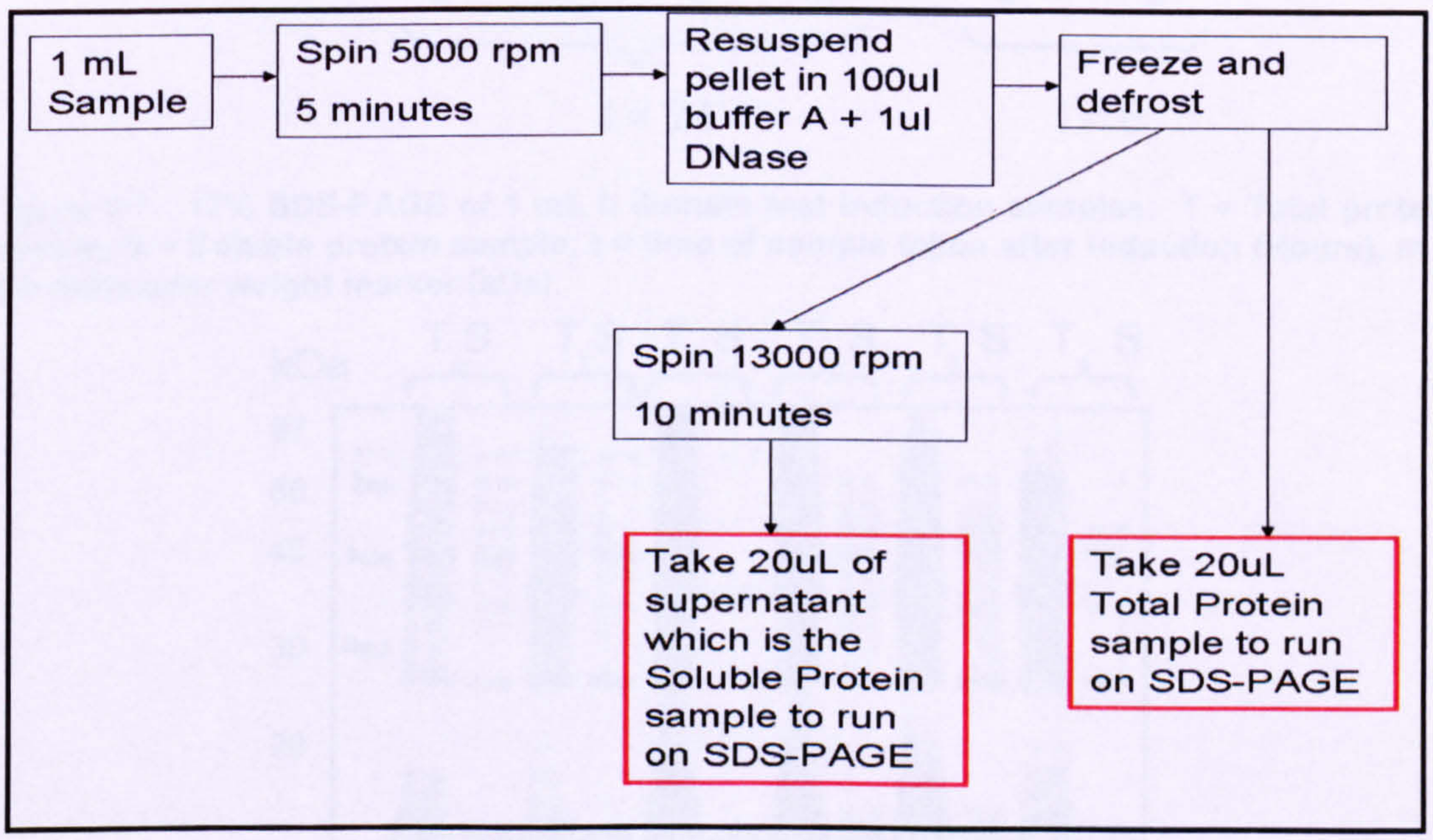

Figure 3-6. Flow diagram of 1 mL sample treatment for SDS-PAGE analysis. (1 µL of 10 mg/l DNase stock was added, samples were centrifuged using a small benchtop centrifuge)

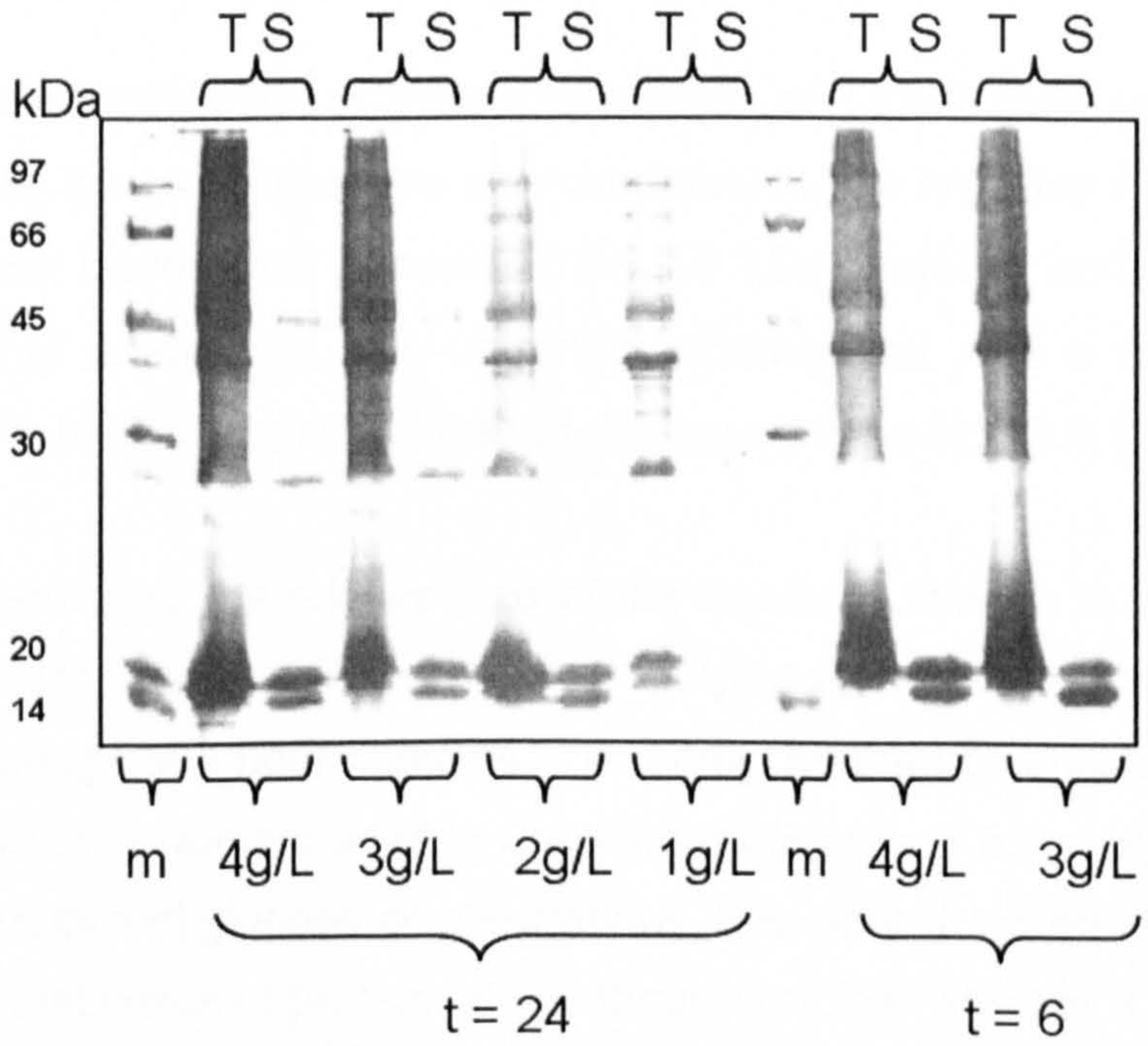

Figure 3-7. 12% SDS-PAGE of 1 mL b domain test induction samples. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), m = low molecular weight marker (kDa).

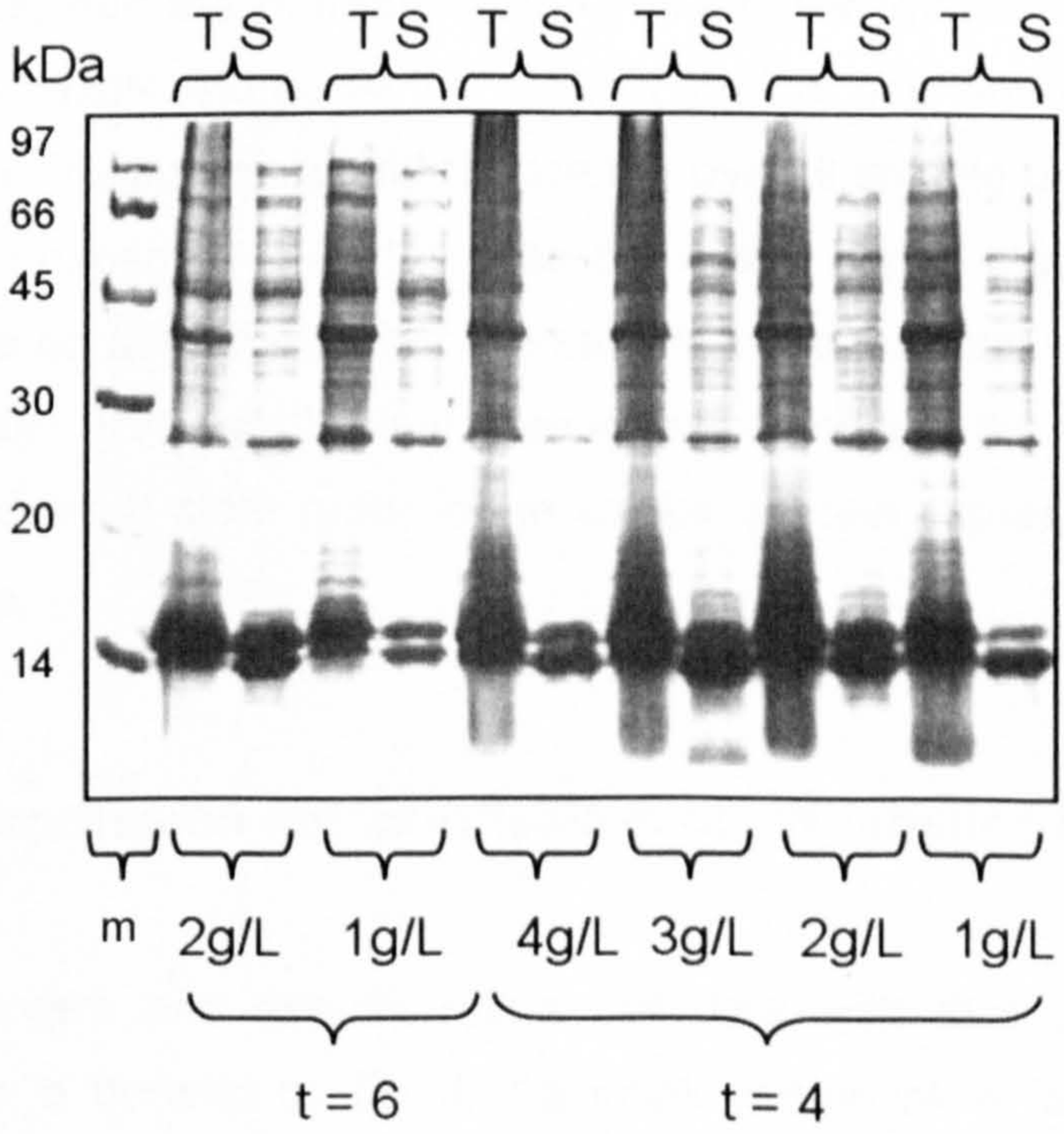

Figure 3-8. 12 % SDS-PAGE of 1 mL b domain test induction samples. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), m = low molecular weight marker (kDa).

The **b** domain protein has a molecular weight of 13147 Daltons and so was believed to be the band that runs approximately in line with the 14 kDa marker. There are two bands that appear at the 14 kDa markers and at this stage (without further analysis) it was uncertain whether this was a soluble protein impurity or the histidine tag modification discussed in section 3.2.1.

Firstly, by comparing the relative size of the bands in relation to induction time (either overnight, 4 or 6 hours); it is clear that the expression of **b** domain is at its highest during the 4 hour samples. Secondly, by then comparing the relative intensities of the **b** domain soluble fractions during the 4 hour expression time and with the different glucose concentrations, it is clear the 3 and 2 g/L glucose gave the highest yields of protein. It is although unclear whether 2 or 3 g/L gave the highest yields. So in the interest of reducing cost of the expensive ^{13}C-glucose reagent, it would be best to use 2g/L when producing ^{13}C-labelled protein samples, half the normal levels of unlabelled glucose used in normal minimal medium expressions. At this stage it has not been investigated whether the reduction in ^{13}C-glucose would reduce the overall soluble protein expression yield. This was expected to be the case but it was hoped that the reduction in yield will not be so detrimental that the sample concentration would be too low for NMR analysis. But what is clear here is that 4 hours is the optimum time for expression as there is clear reduction in soluble protein expression beyond this expression time.

3.2.3. Expression and purification of ^{15}N-labelled b domain

In order to confirm and directly relate our data with that of the published structure of the **b** domain by Dr. J. Kemmink (Kemmink *et al.*, 1999), it was essential to reproduce and assign the **b** domain independently. The conditions at which we aim to collect our NMR data differ from those used in the published

b domain and so we need to identify any differences in the spectra that may arise. It was not possible to use the spectra and assignments deposited by Dr. Kemmink as slight difference in preparation or buffers could affect the residue shifts seen on a HSQC spectrum. In order to first assess any difference in preparation and assess the quality of the data, it was required to produce a ^{15}N-labelled **b** domain protein sample, from which we can collect a 2D-HSQC spectrum.

The **b** domain was expressed as described in section 2.2.7 in minimal medium with ^{15}N-labelled ammonium sulphate, the preparation of which is described in section 2.2.10. The purification procedure was the same as all expression purifications, starting with IMAC as described in section 2.3.1 and the SDS-PAGE analysis of which is shown in Figure 3-9. The IMAC elute was then further purified using anion-exchange chromatography as described in section 2.3.2. 20 µL samples from each of the 2 mL peak fractions were treated as described in section 2.2.11 and run on an SDS-PAGE gel shown in Figure 3-9 and Figure 3-10.

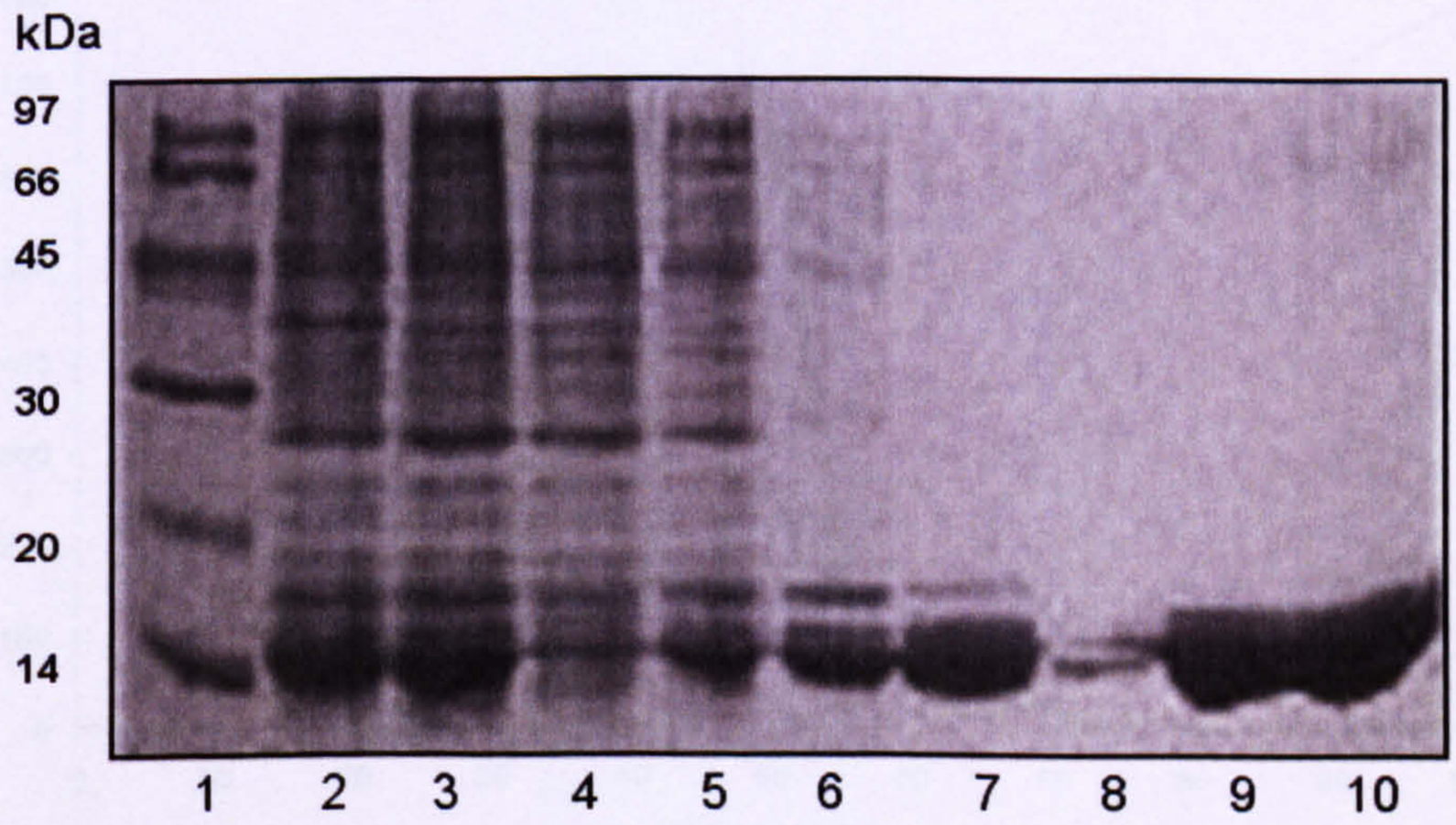

Figure 3-9. 16% SDS-PAGE of ^{15}N b domain applied to IMAC and Anion-exchange fractions. Lane 1, low molecular weight marker. Lane 2, total protein. Lane 3, soluble protein. (Lanes 4-7; analysis of IMAC purification) Lane 4, flow through. Lane 5, imidazole wash. Lane 6, low salt wash. Lane 7, hexa-histidine tagged protein elute. Lane 8-10 fractions from anion-exchange Source 30Q column.

It is clear that IMAC purification step was very effective at removing almost all of non-tagged protein impurities, except one band of a slightly larger molecular weight. To ensure all impurities were removed, the elute was then prepared and further purified on an anion-exchange column as described in section 2.3.2.

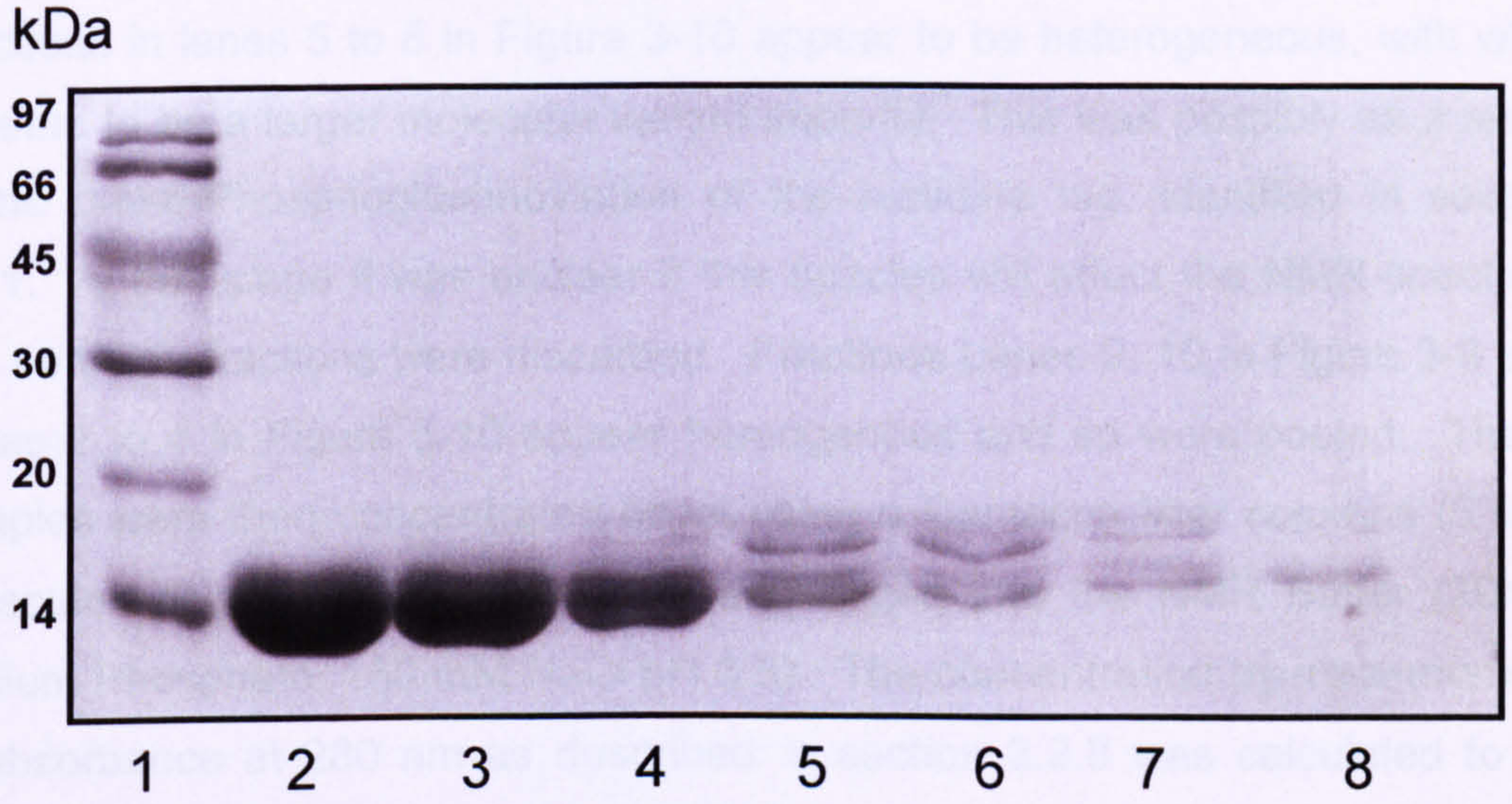

Figure 3-10. 16% SDS-PAGE of anion-exchange fractions of ^{15}N-labelled b domain. Lane 1, low molecular weight marker. Lane 2 – 8 fractions from across the peak in the anion-exchange Source 30Q column.

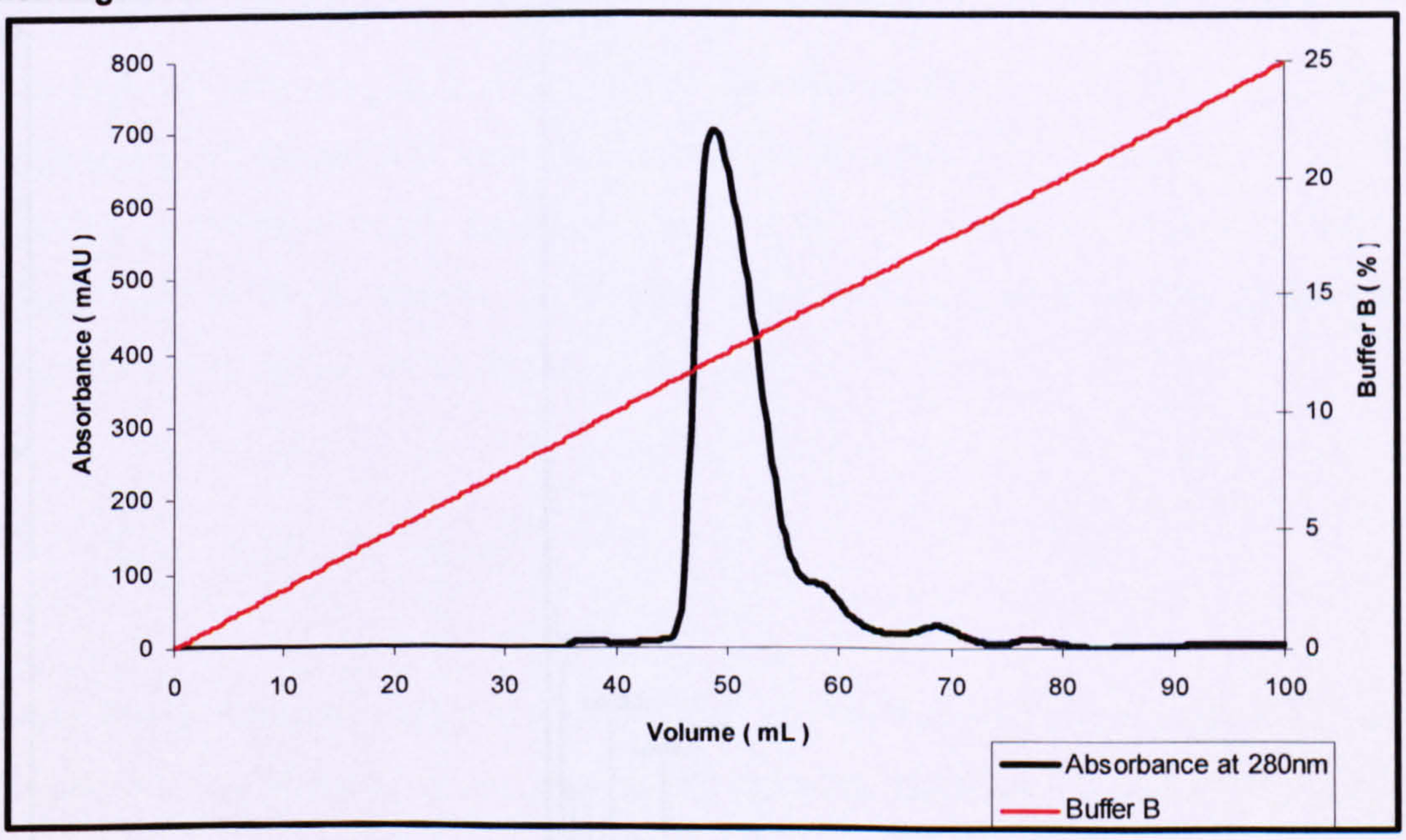

Figure 3-11. Chromatogram of ^{15}N b domain using a Source 30Q column, eluted with a 25% Buffer B gradient over 100mL.

From the anion-exchange chromatogram in Figure 3-11, the eluted **b** domain appears to be homogenous, eluting in one peak, with a slight shoulder at end. To analyse the homogeneity of the protein, samples were taken from each of the 2mL fractions and run on SDS PAGE see Figure 3-9 and Figure 3-10.

Fractions in lanes 5 to 8 in Figure 3-10 appear to be heterogeneous, with what appears to be a larger molecular weight impurity. This was possibly as a result of the α-N-6-Phosphogluconoylation of the histidine tag, identified in section 3.2.1. At this stage it was unclear if this species will affect the NMR spectrum and so these fractions were discarded. Fractions Lanes 9, 10 in Figure 3-9 and Lanes 2 to 4 in Figure 3-10 appear homogenous and so were pooled. These samples were then concentrated down using a Centricon filter columns (5 kDa molecular weight cut off) and buffer exchanged into the NMR Buffer (20mM Sodium Phosphate, 150 mM NaCl pH 6.5). The concentration by measurement of absorbance at 280 nm as described in section 2.2.8 was calculated to be 28.2 mg, 2.12 mM in a 0.5 mL solution.

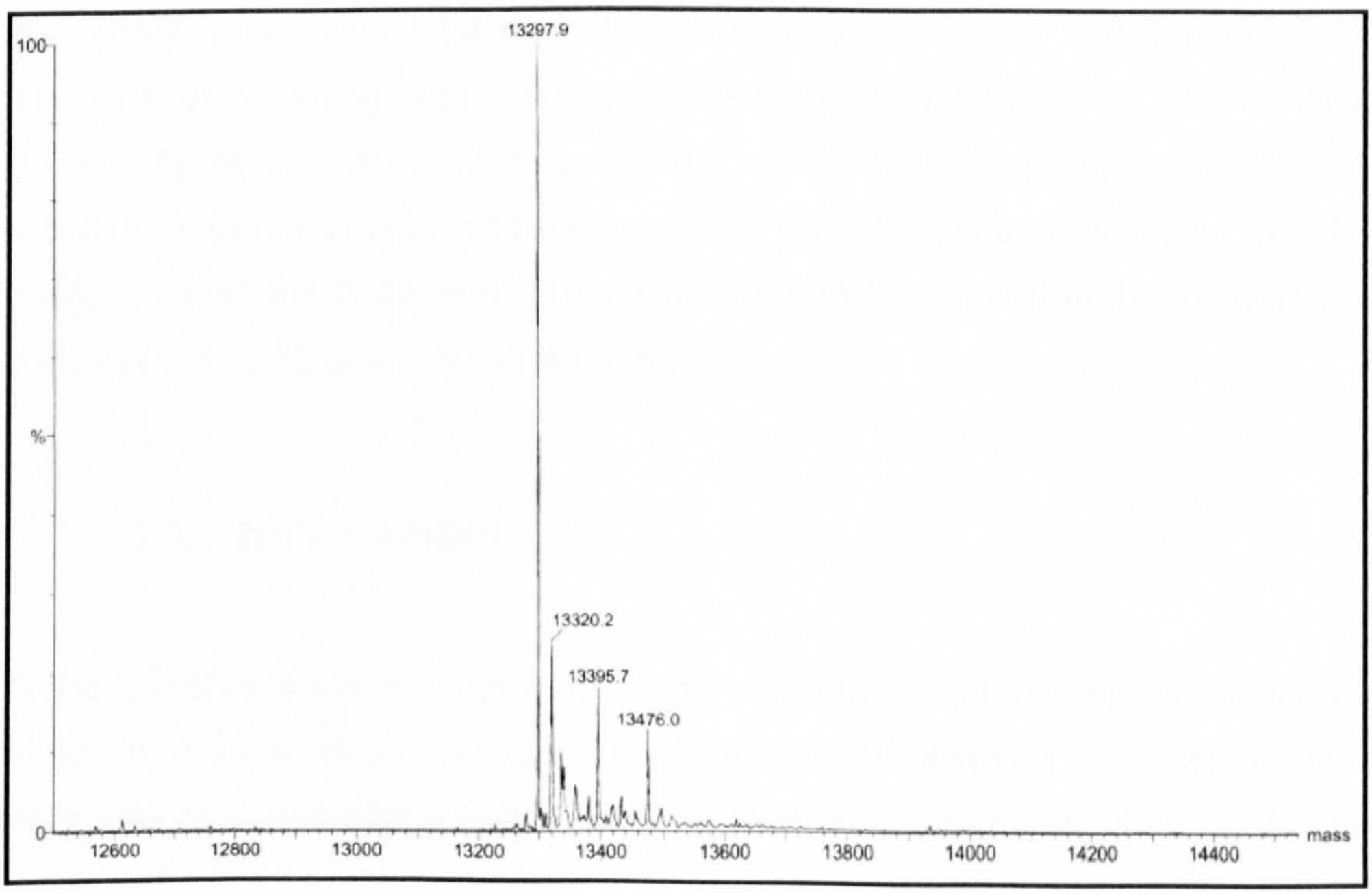

Figure 3-12. Positive ion mass spectrum of the ^{15}N b domain.

The expected molecular weight if all 153 nitrogen atoms were labelled with the ^{15}N isotope would be 13300 Da. The actual molecular weight of the major species recorded by electrospray mass spectrometry was 13298 Da, which indicates 2 of the 153 nitrogen's were not labelled and so the labelling efficiency was 98.7%.

The mass spectrum in Figure 3-12 also reveals a sodium ion adduct at 13320.2 Da and the dephosphorylated α-N-6-Phosphogluconoylation of the histidine-tag modification at 13476 Da (+178 Da). There was also an unknown adduct with an extra mass of 98 Da, with a total mass of 13395.7 Da, this could be as a result of a phosphorylation (H_2O_3PO-) of the protein at an unknown site

This initial work established the optimum expression conditions for the **b** domain. The ^{15}N-labelled **b** domain sample described here gave good spectra to allow assignment and indeed as expected there were subtle differences in chemical shifts resulting from differing sample conditions to the published **b** domain data, this analysis is shown and described further in Figure 5-8 and section 5.3.3. The task of obtaining **b** domain assignments was taken over by Dr. K. Wallis (University of Warwick), who proceeded successfully to produce a ^{13}C, ^{15}N-labelled **b** domain NMR sample, using 2 g/L ^{13}C-glucose, enabling the full assignment of the **b** domain. These assignments are used in this research to assist in the assignment of the **bb'x** domain.

3.3. bb'x domain

Since the **b'** domain was the main target of interest and since prior NMR data shows that the **b'** alone gives poor NMR spectra, **bb'x** was a better target since **bb'x** was shown to give a better resolved NMR spectra as described in Chapter

4. So production of high yields and homogenous isotopically labelled **bb'x** samples for NMR analysis are crucial to enable the collection of good NMR data.

3.3.1. Expression and characterisation of the bb'x domain

The **bb'x** domain construct was grown as described in section 2.2.7. Two 400 mL minimal medium cultures were grown in two 2 Litre flasks. The sample was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1.

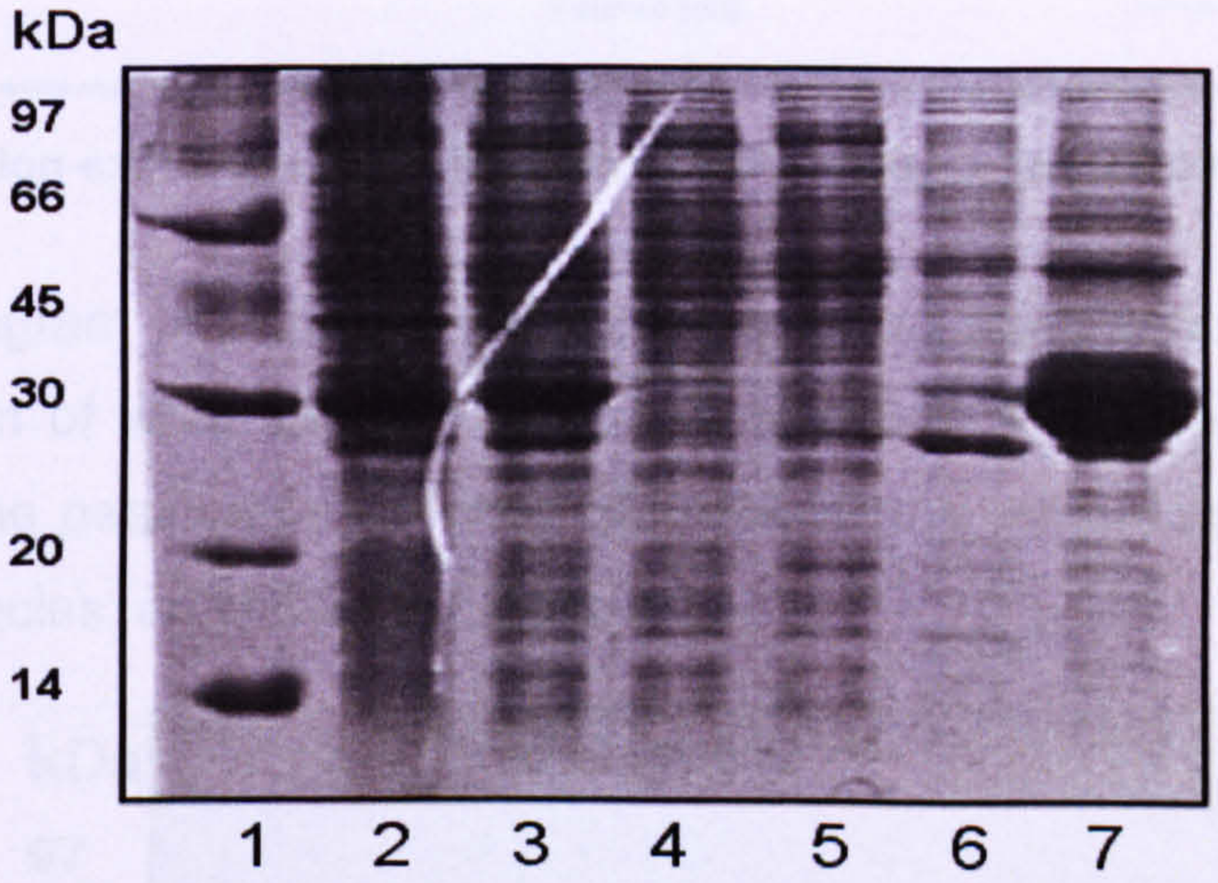

Figure 3-13. 16% SDS-PAGE of bb'x applied to IMAC. Lane 1, low molecular weight marker. Lane 2, total protein. Lane 3, soluble protein. Lane 4, flow through. Lane 5, imidazole wash. Lane 6, low salt wash. Lane 7, hexa-histidine tagged bb'x protein elute.

The IMAC purification, as shown in Figure 3-13, was not as efficient as seen previously, this may result of the high level of **bb'x** protein expression; hence overloading of the column. The IMAC elute was therefore further purified using anion-exchange chromatography as described in section 2.3.2.

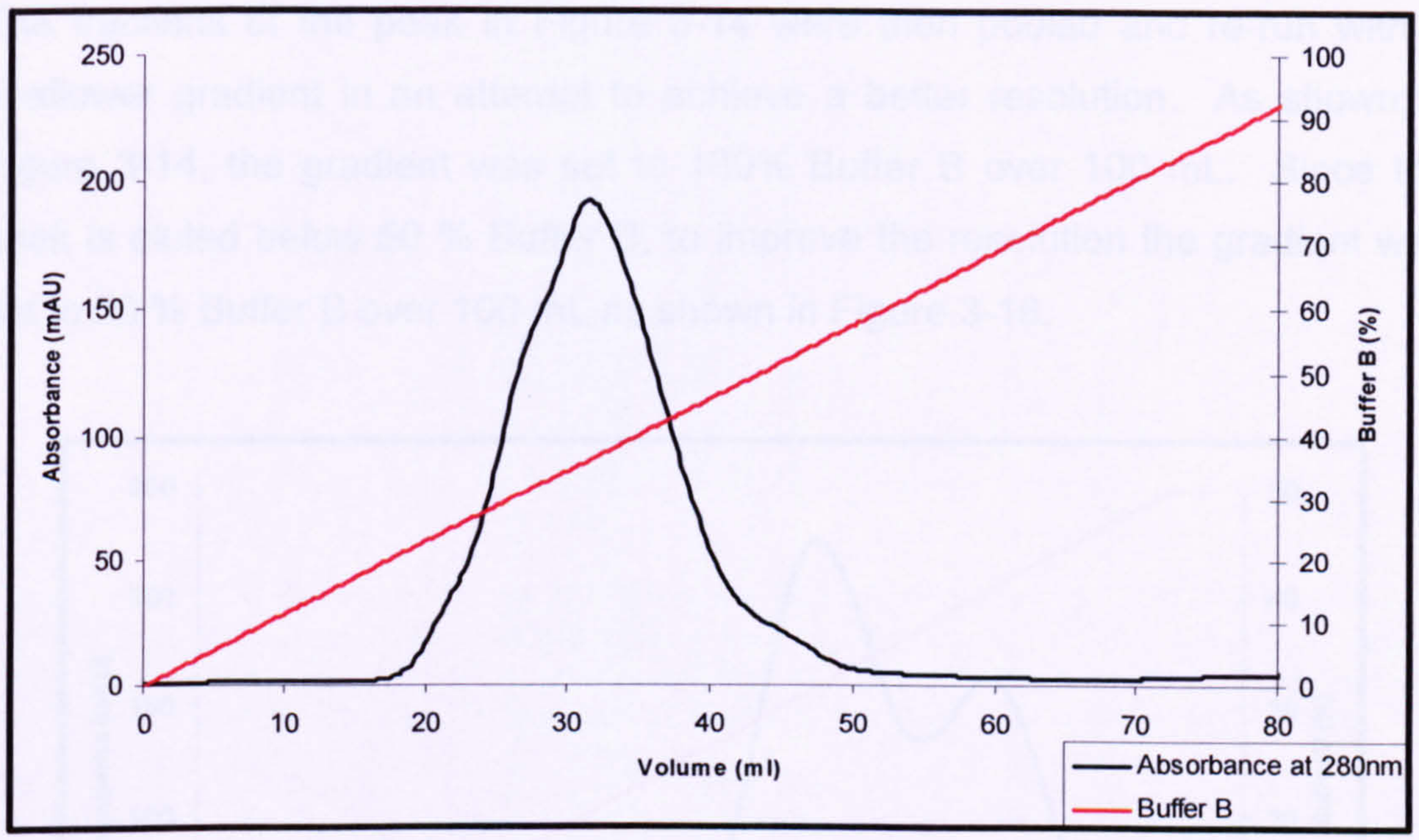

Figure 3-14. Anion-exchange chromatogram of bb'x using a Source 30Q column.

The chromatogram in Figure 3-14 shows a broad peak which results from a poor resolution of what appears to be two species. But when samples taken from across the peak were run on a SDS-PAGE the protein appears to consist of a single species, as shown in Figure 3-15.

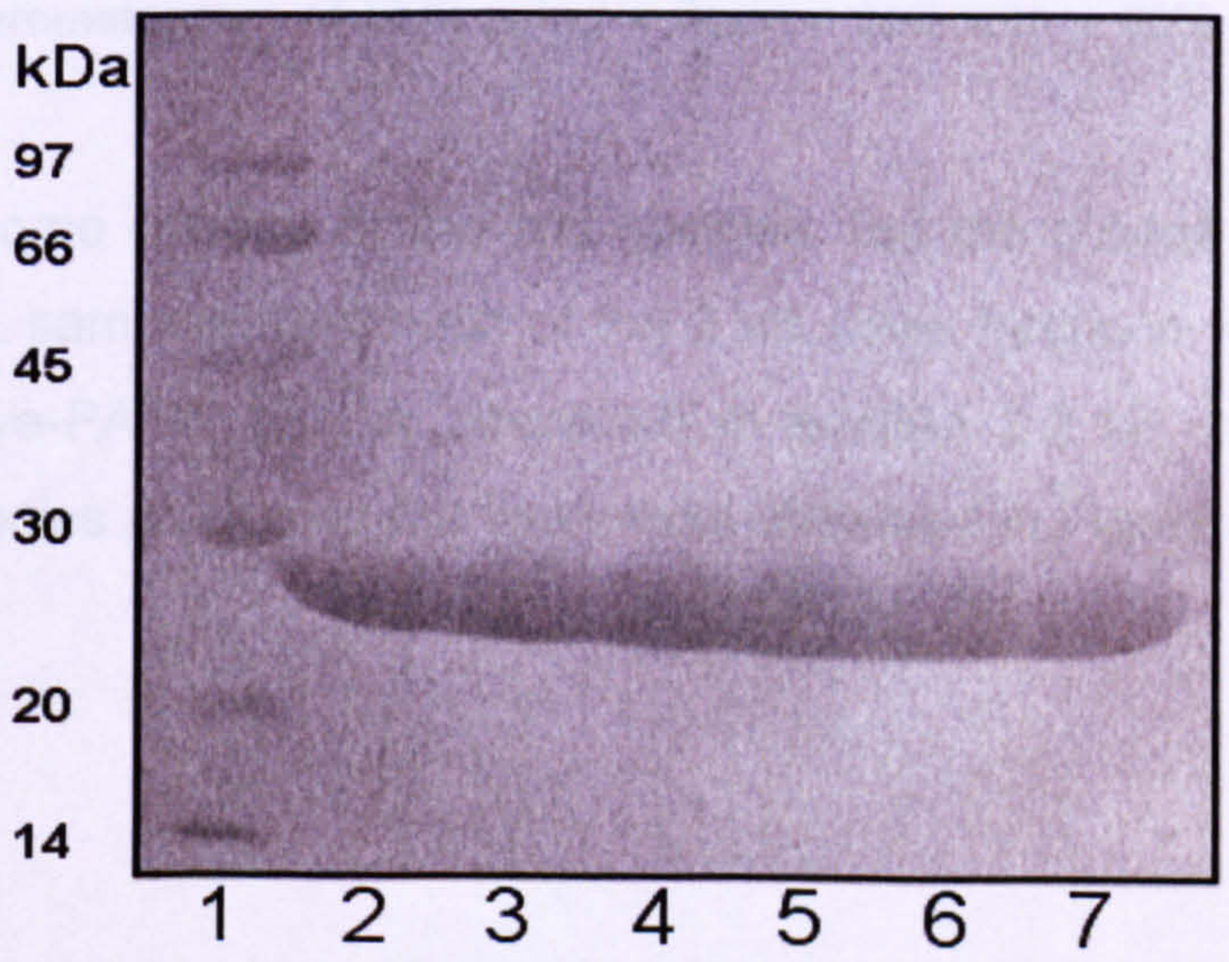

Figure 3-15. 16% SDS-PAGE of fractions taken across the peak in chromatogram peak shown in Figure 3-14.

The fractions of the peak in Figure 3-14 were then pooled and re-run with a shallower gradient in an attempt to achieve a better resolution. As shown in Figure 3-14, the gradient was set to 100% Buffer B over 100 mL. Since the peak is eluted below 50 % Buffer B, to improve the resolution the gradient was set to 50 % Buffer B over 100 mL as shown in Figure 3-16.

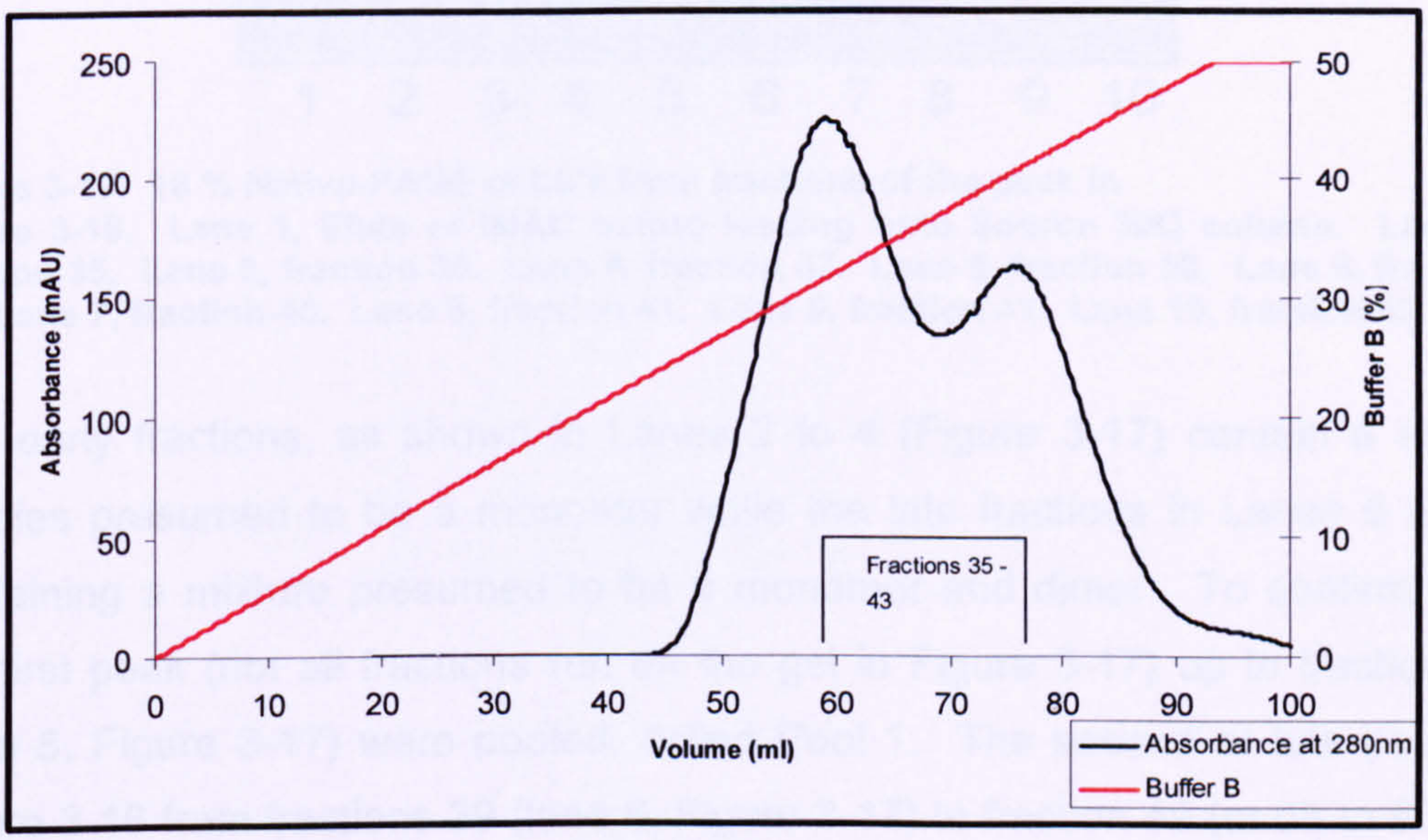

Figure 3-16. Chromatogram of bb'x using a Source 30Q with a 50% Buffer B gradient over 100mL

There is still some overlap of the two species, but the overall ratio was much smaller. 20 µL samples from each of the 2 mL peak fractions were treated and run on a native-PAGE gel as described in section 2.2.12; in an attempt to identify the species producing the two peaks observed in Figure 3-16.

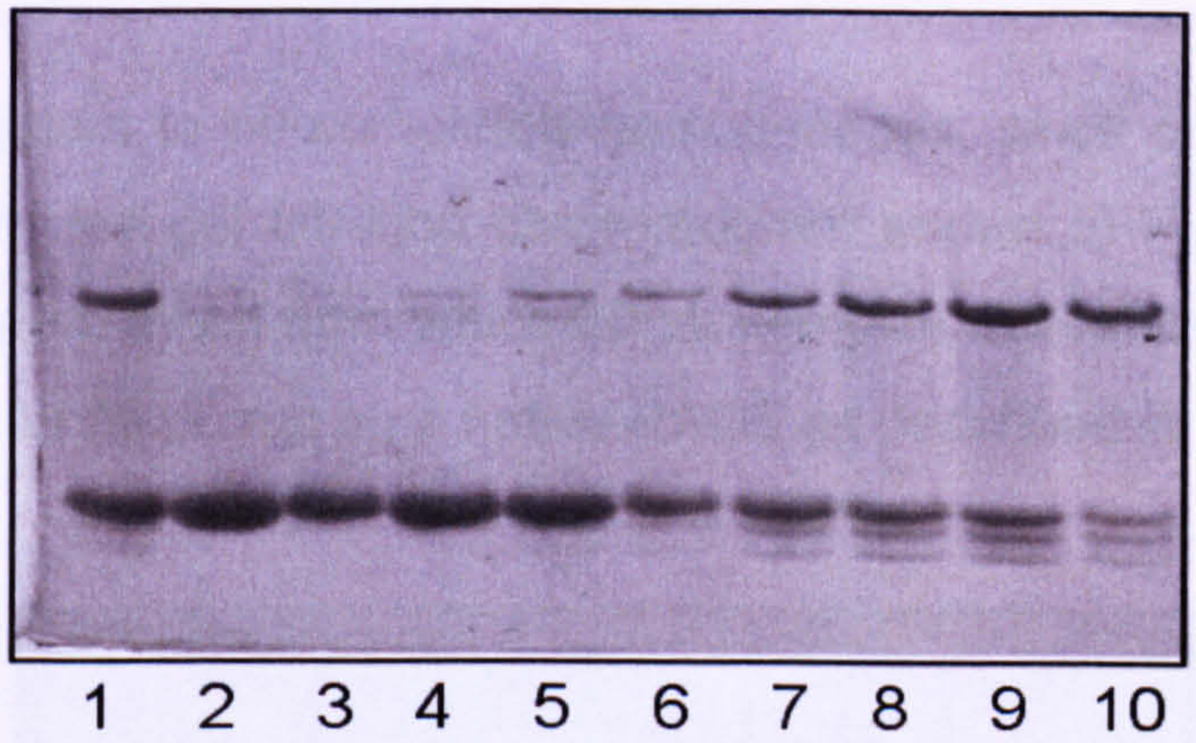

Figure 3-17. 16 % Native-PAGE of bb'x from fractions of the peak in Figure 3-16. Lane 1, Elute of IMAC before loading onto Source 30Q column. Lane 2, fraction 35. Lane 3, fraction 36. Lane 4, fraction 37. Lane 5, fraction 38. Lane 6, fraction 39. Lane 7, fraction 40. Lane 8, fraction 41. Lane 9, fraction 42. Lane 10, fraction 43.

The early fractions, as shown in Lanes 2 to 4 (Figure 3-17) contain a single species presumed to be a monomer while the late fractions in Lanes 5 to 10 containing a mixture presumed to be a monomer and dimer. To confirm this, the first peak (not all fractions run on the gel in Figure 3-17) up to fraction 38 (lane 5, Figure 3-17) were pooled, called Pool 1. The second or late peak in Figure 3-16 from fractions 39 (lane 6, Figure 3-17) to fraction 49 (or 68 to 90 mL) (not all fractions run on gel in Figure 3-17) were pooled separately, called Pool 2. Each pool was then applied to a Superdex 75 Gel filtration column as described in section 2.3.3.

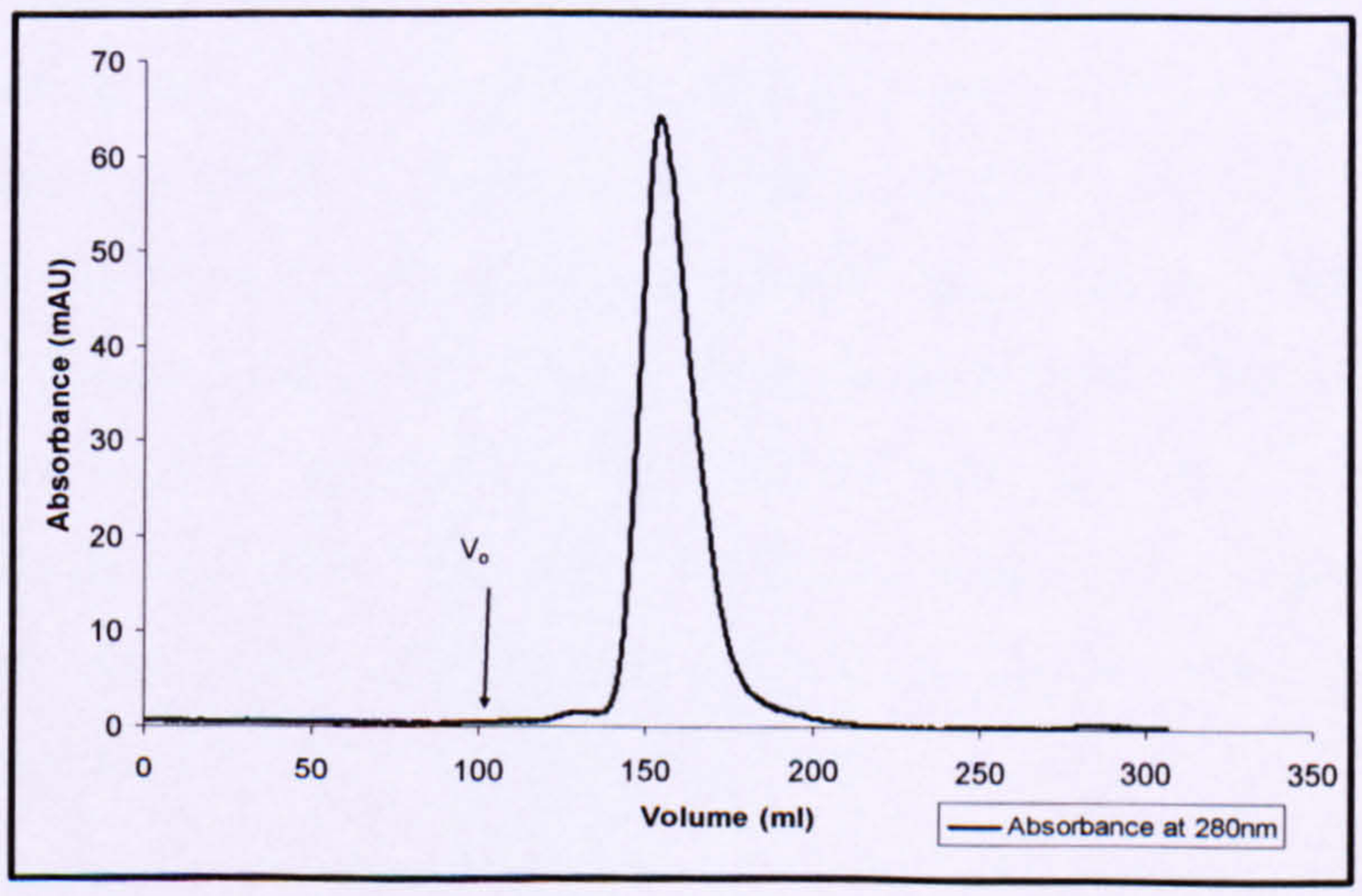

Figure 3-18. Superdex 75 gel filtration chromatogram of Pool 1. The void volume highlighted with V_o was 105 mL.

Pool 1 now appears to be completely homogeneous, since only a single peak was visible from the gel filtration chromatogram shown in Figure 3-18, 3 mL peak fractions were taken from which 20 µL samples where treated with native-loading buffer and 20 µL run on a native-PAGE as described in section 2.2.12.

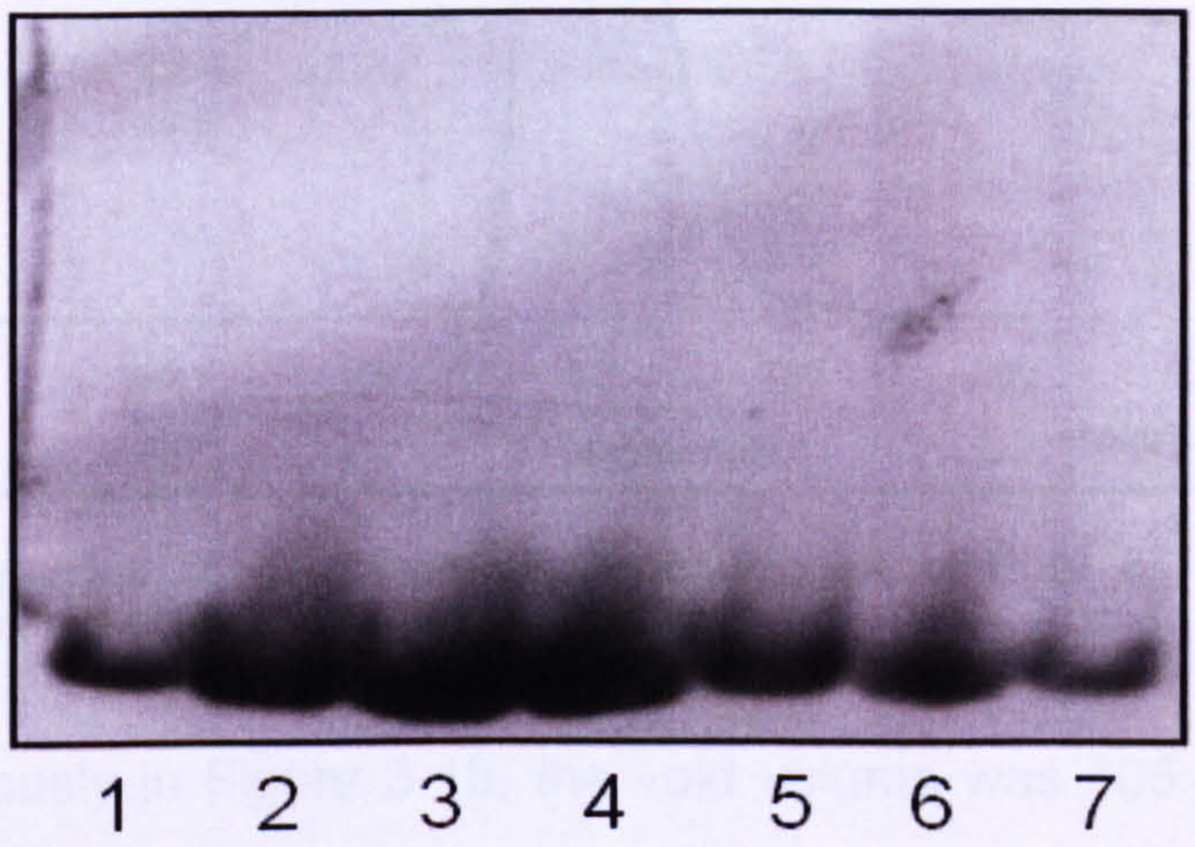

Figure 3-19. 16 % native-PAGE of fractions taken across peak in chromatogram in Figure 3-18.

The fractions across the peak in Figure 3-18 are in fact homogenous as shown in Figure 3-19, whereby no higher molecular weight protein bands are visible. All fractions from the peak in Figure 3-18 were pooled and called monomer.

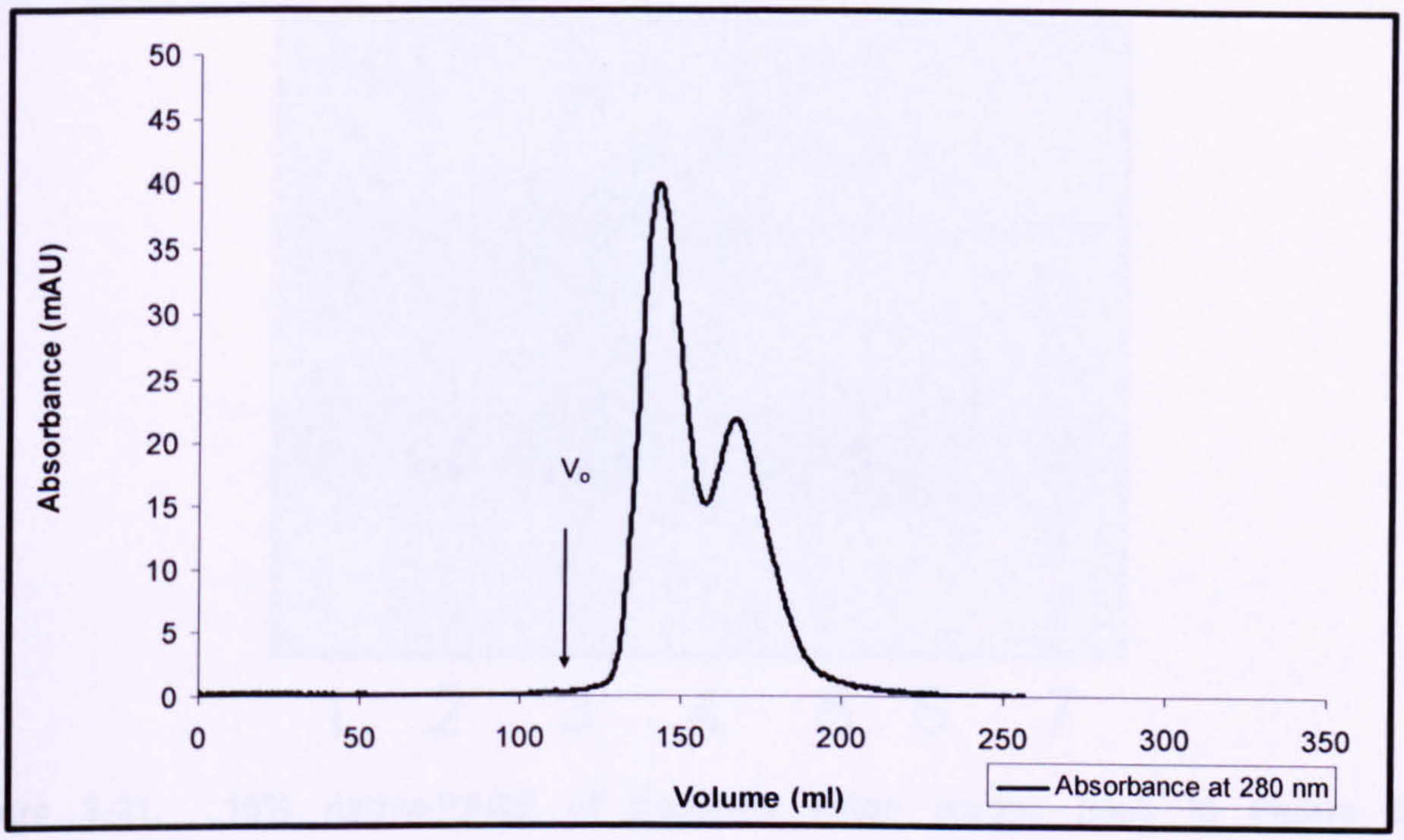

Figure 3-20. Superdex 75 gel filtration chromatogram of Pool 2. The void volume highlighted with V_o is 115mL

Note that previously in Figure 3-18, the void volume was 105 mL and the first peak elutes approximately 55 mL after the void volume at approximately 160 mL (accumulated volume). The Pool 2 gel filtration column in Figure 3-20 was started 10 mL late (void volume was therefore 115 mL), so the same species which eluted at 160 mL (in Figure 3-18) now elutes at approximately 170 mL (later peak). The larger molecular weight species, which was first identified in the anion-exchange column run in Figure 3-17 and seen on native gel in Figure 3-17, is therefore the first or early peak eluted in the gel filtration column 25 mL after the void volume of 115mL at approximately 140 mL. From the chromatogram in Figure 3-20, it appears that Pool 2 was a mixture of the two species, as expected since the Pool 2 fractions appeared to have both low and high molecular weight species in the native-PAGE analysis in Figure 3-17. Therefore, 20 µL samples from the 3 mL peak fractions in Figure 3-20 were taken treated with loading buffer and 20 µL run on a native-PAGE gel as described in section 2.2.12.

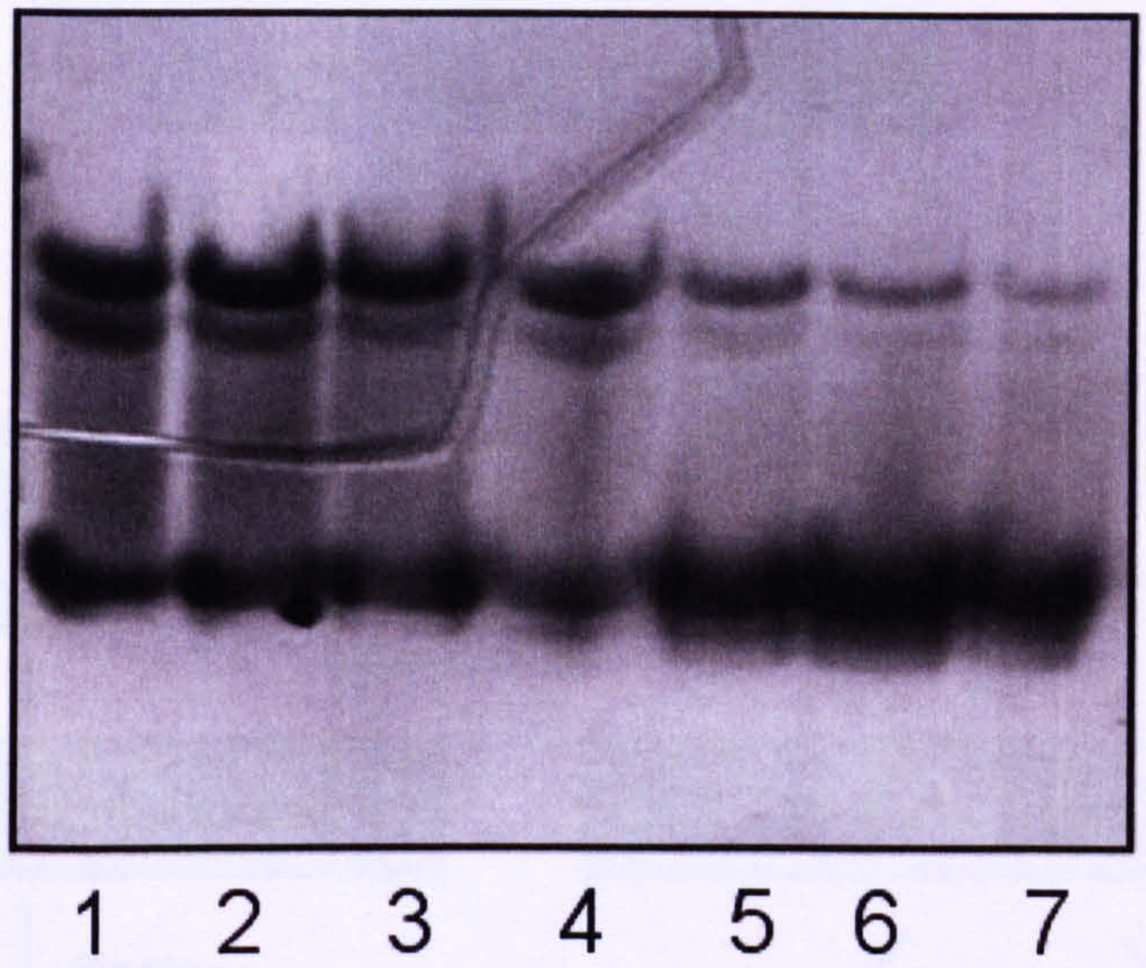

Figure 3-21. 16% native-PAGE of fractions taken across peak in Figure 3-20 chromatogram.

From Figure 3-21, the native gel analysis of the chromatogram in Figure 3-20, reveals two species, the larger molecular weight peak (dimer) predominately residing in the first peak from the gel filtration chromatogram in Figure 3-20 and the smaller molecular weight species (monomer) residing in the later peak of the gel filtration chromatogram. From the gel therefore, fractions in lanes 1 to 4 were pooled together into dimer pool, while fractions in Lanes 5 and 7 were added to the previously described monomer pool in Figure 3-19.

As mentioned earlier and now somewhat clearer, pooled fractions from anion-exchange were titled Pool 1, early peak and Pool 2 for the later peak. These samples further purified and analysed by native-PAGE revealed Pool 1 samples constituted largely of a monomer species and the Pool 2 samples were somewhat heterogeneous but could be further fractionated to mainly consist of the dimer species. This is summarised in Figure 3-22.

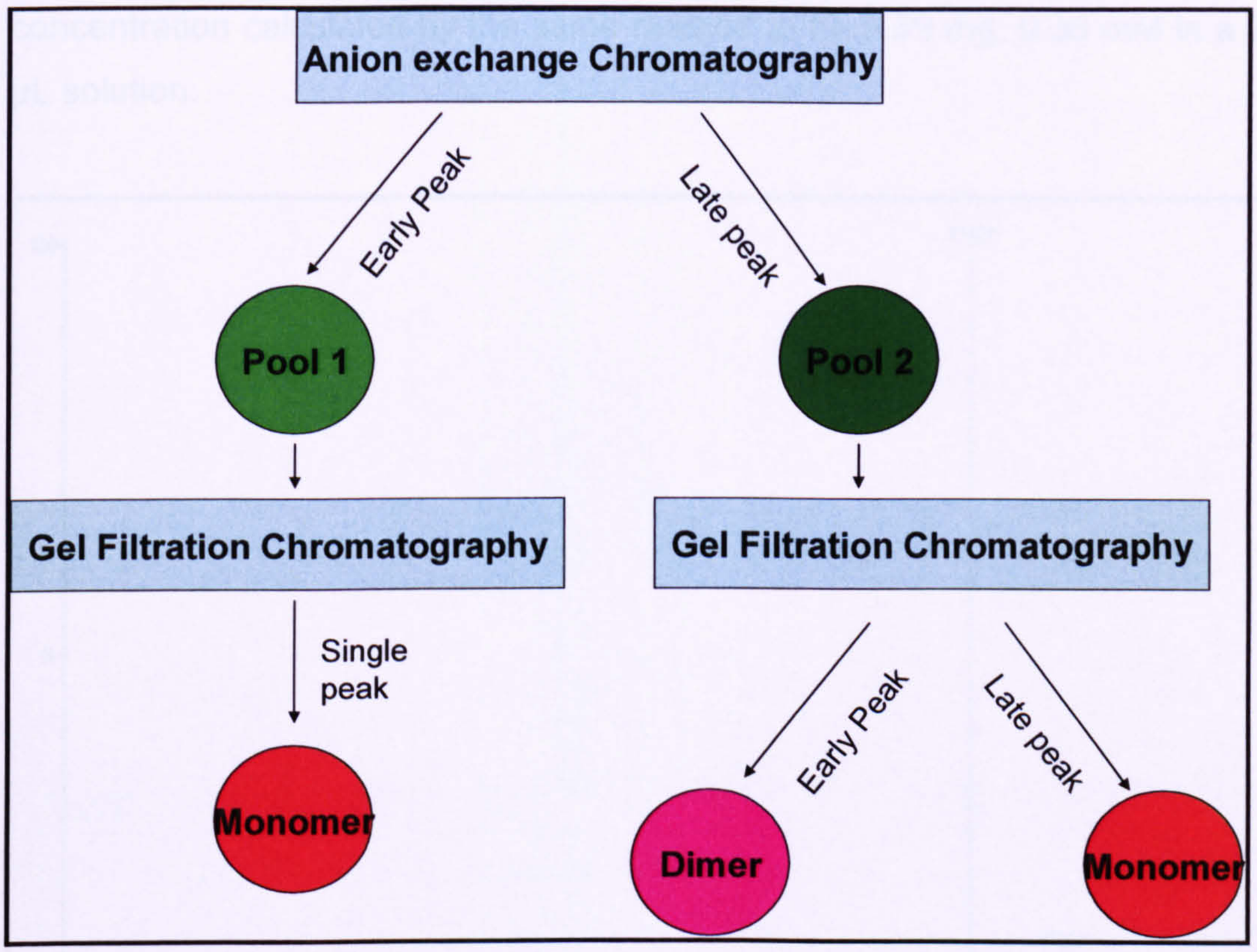

Figure 3-22. Diagram showing the fractionation and pool designation. Note that fractionation of the early peak after anion-exchange generally is sufficient to isolate the monomer species, the dimer is rarely seen in Pool 1 treatment by gel-filtration.

It was at this stage still an assumption that the low and high molecular weight species are monomer and dimer, respectively. But from the SDS-PAGE analysis, as shown in Figure 3-15 it appears that the protein is a single homogenous species. The evidence of the presence of the higher molecular weight species in Pool 2 samples comes from the fact that two species are seen on native-PAGE analysis and can be separated by gel filtration. Confirmation that this species is a dimer was carried out and is discussed later in Chapter 3.8.

The monomer pool was concentrated down using Centricon filter columns (5 kDa molecular weight cut off) and the concentration was measured using the Biorad assay as described in section 2.2.8. The concentration was calculated to be 98.9 mg, 3.6 mM in a 200 µL solution. The dimer pool was concentrated and

concentration calculated by the same method to be 9.89 mg, 0.36 mM in a 250 µL solution.

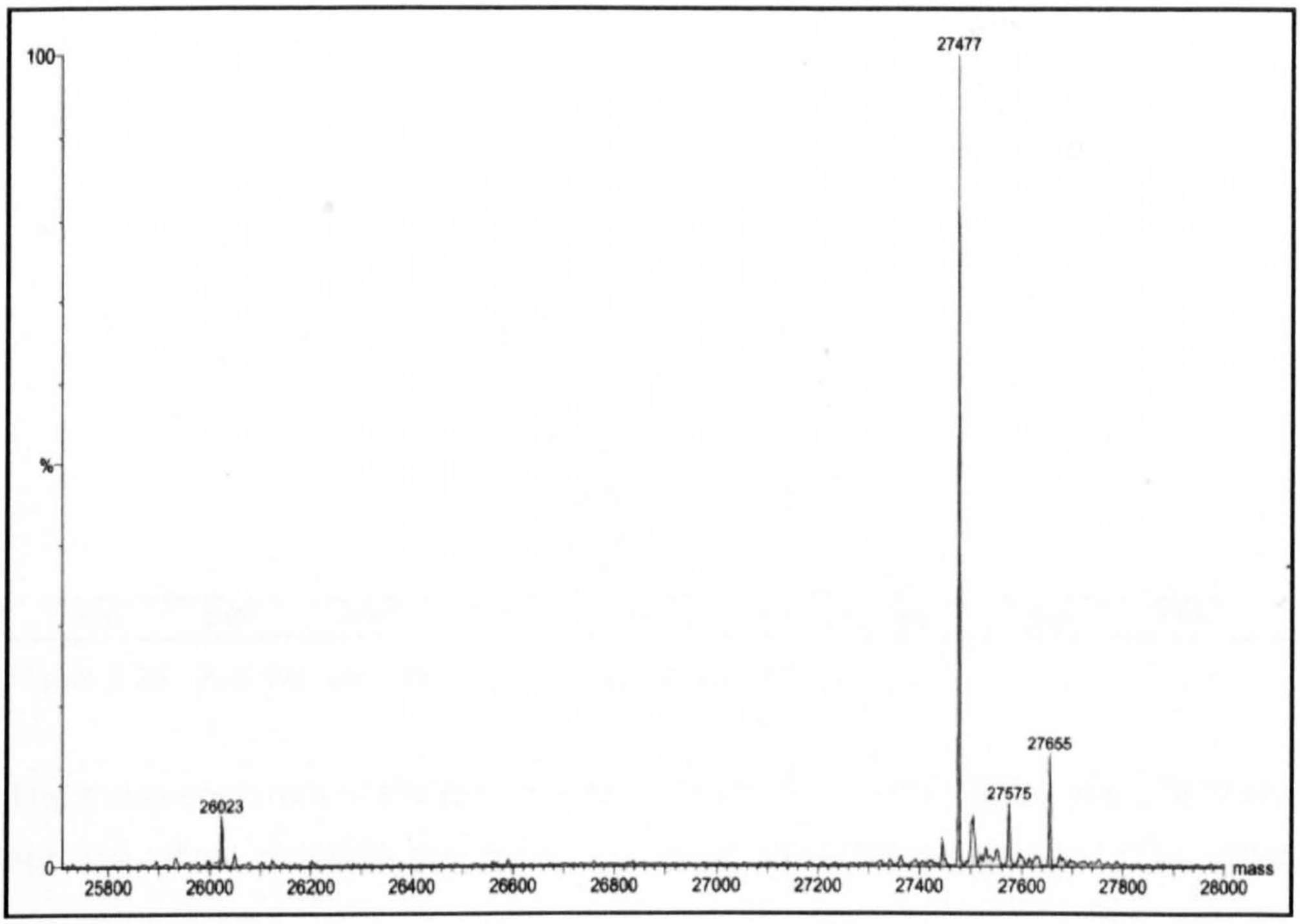

Figure 3-23. Positive ion mass spectrum of the bb'x monomer.

The mass spectrum of the **bb'x** monomer in Figure 3-23 reveals the major species has mass of 27477 Da which matches the predicted mass from the sequence. It also shows the dephosphorylated (+178 Da) α-N-6-Phosphogluconoylation of the histidine-tag at 27655 Da. There was also an unknown adduct, with an extra mass of 98 Da, with a total mass of 27575 Da, this could be as a result of a phosphorylation (H_2O_3PO-) of the protein at an unknown site. There also appears to be a very minor species possibly a degradation product with a -1454 Da with a total mass of 26023 Da.

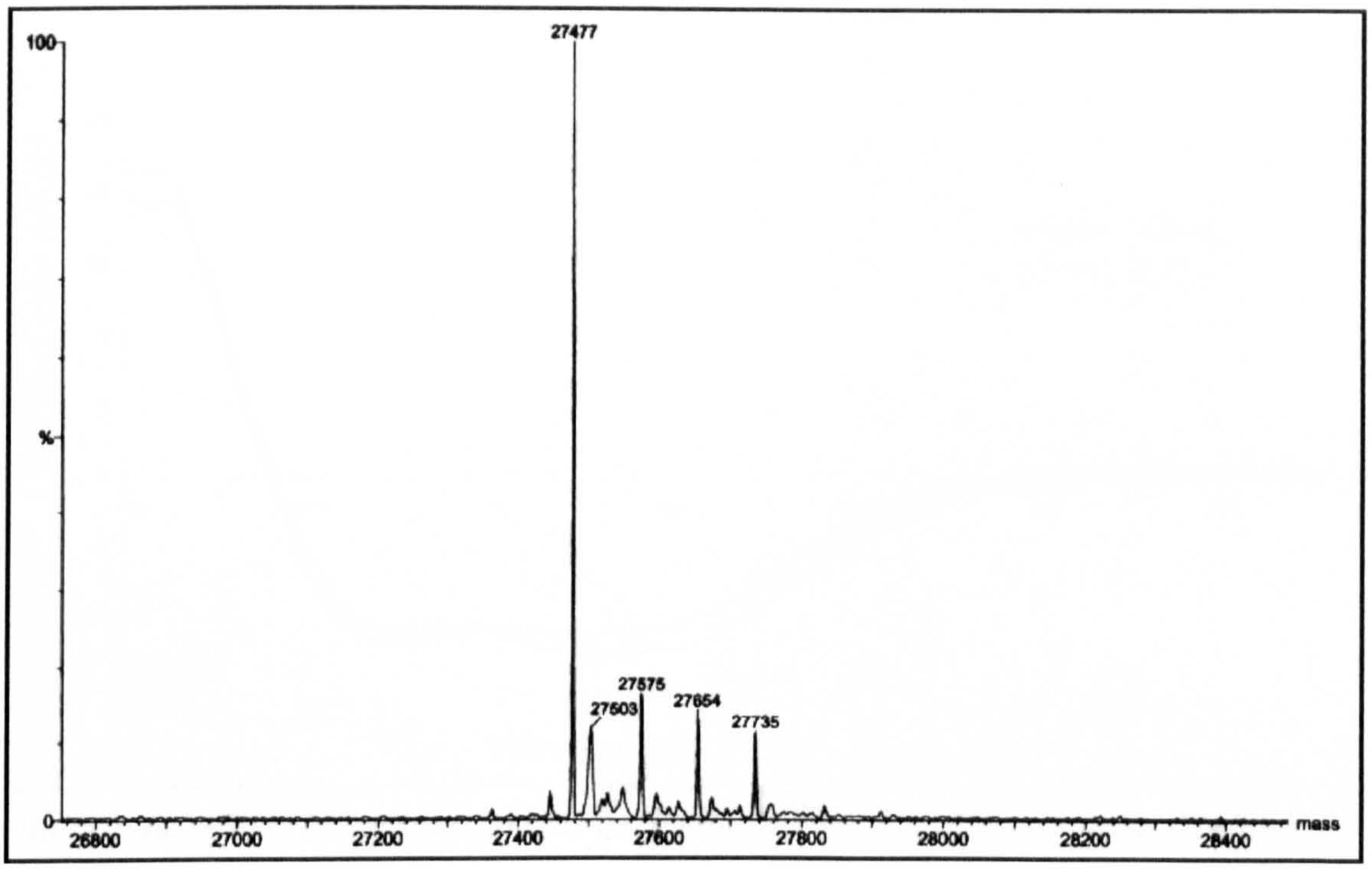

Figure 3-24. Positive ion mass spectrum of the bb'x dimer.

The mass spectrum of the **bb'x** dimer in Figure 3-24 also reveals the 27477 Da species which matches the predicted mass using the sequence. The other species are very minor; at approximately 10% intensity of the major species observed and are the same as those observed previously.

There appears to be no difference in mass spectrometry data between the monomer and dimer species, importantly no dimer mass was recorded; it was likely that the treatment during the mass spectrometry preparation and analysis disrupts the dimer and as a result no dimer species was detected.

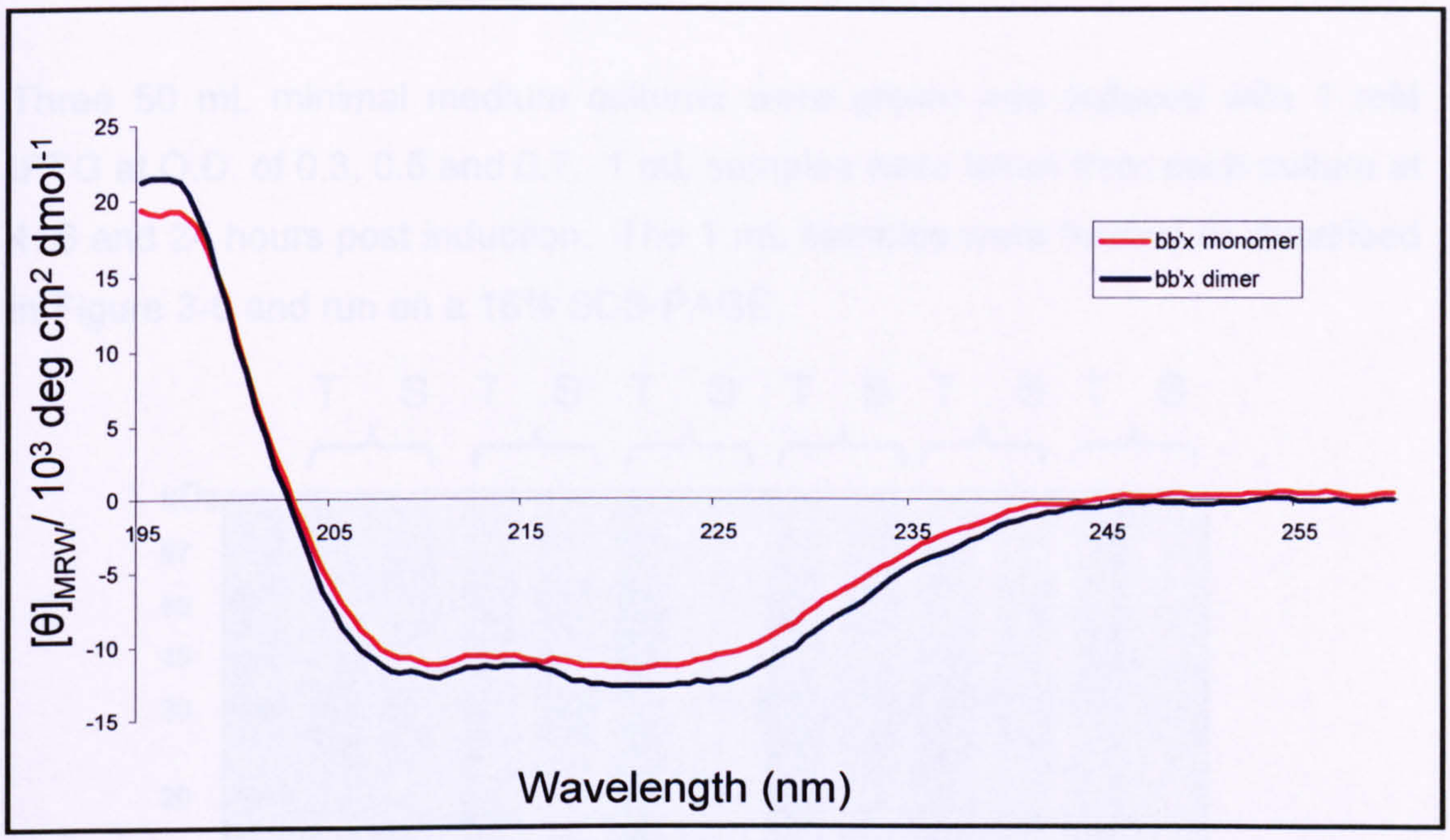

Figure 3-25. A Far UV CD spectrum for the two pools of bb'x domain. Using a sample, of 2.5uM, 0.07 mg/ml.

From the CD spectrum shown in Figure 3-25, it is clear that both recombinantly expressed **bb'x** domain monomer and dimer are folded and there was no major structural difference between the two species samples. Analysis of the spectrum indicated 39% and 15% helical and beta content, respectively, for the **bb'x** monomer sample. Also, 44% and 15% helical and beta content, respectively, were recorded for **bb'x** dimer sample.

3.3.2. Optimisation of bb'x expression

The aim of these experiments was to identify the optimum conditions in terms the cell density before induction with IPTG and protein expression time once induced with IPTG. Also, in light of producing ^{13}C-labelled protein for NMR experiments, the aim was to also identify the optimum induction time in relation to reduced amount of glucose used. Isotopically labelled ^{13}C-glucose was an expensive reagent and so reducing the quantities used was financially beneficial.

Three 50 mL minimal medium cultures were grown and induced with 1 mM IPTG at O.D. of 0.3, 0.5 and 0.7. 1 mL samples were taken from each culture at 4, 6 and 24 hours post induction. The 1 mL samples were treated as described in Figure 3-6 and run on a 16% SDS-PAGE.

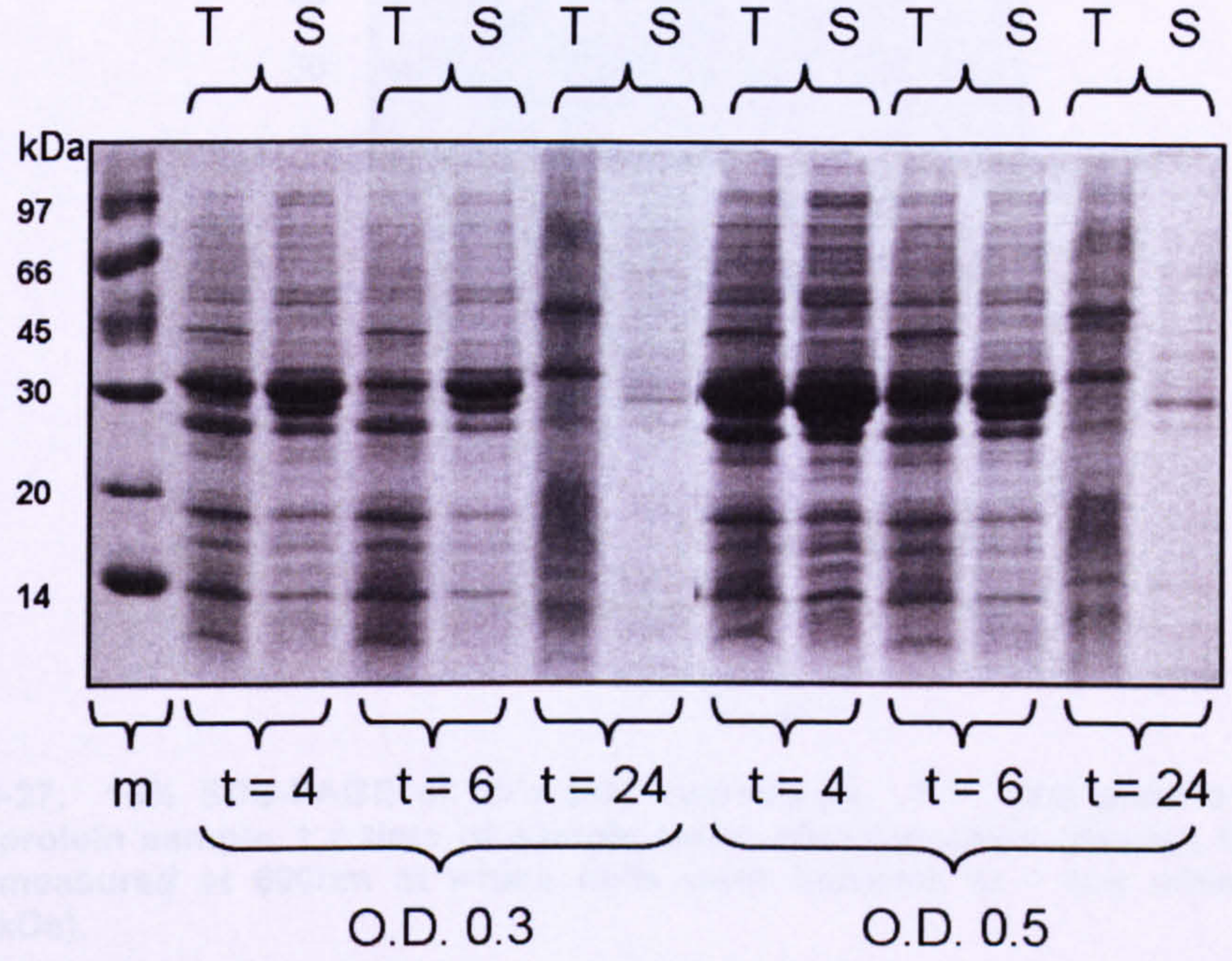

Figure 3-26. 16 % SDS-PAGE of bb'x test expression. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), O.D. = optical density measured at 600nm at which cells were induced, m = low molecular weight marker (kDa).

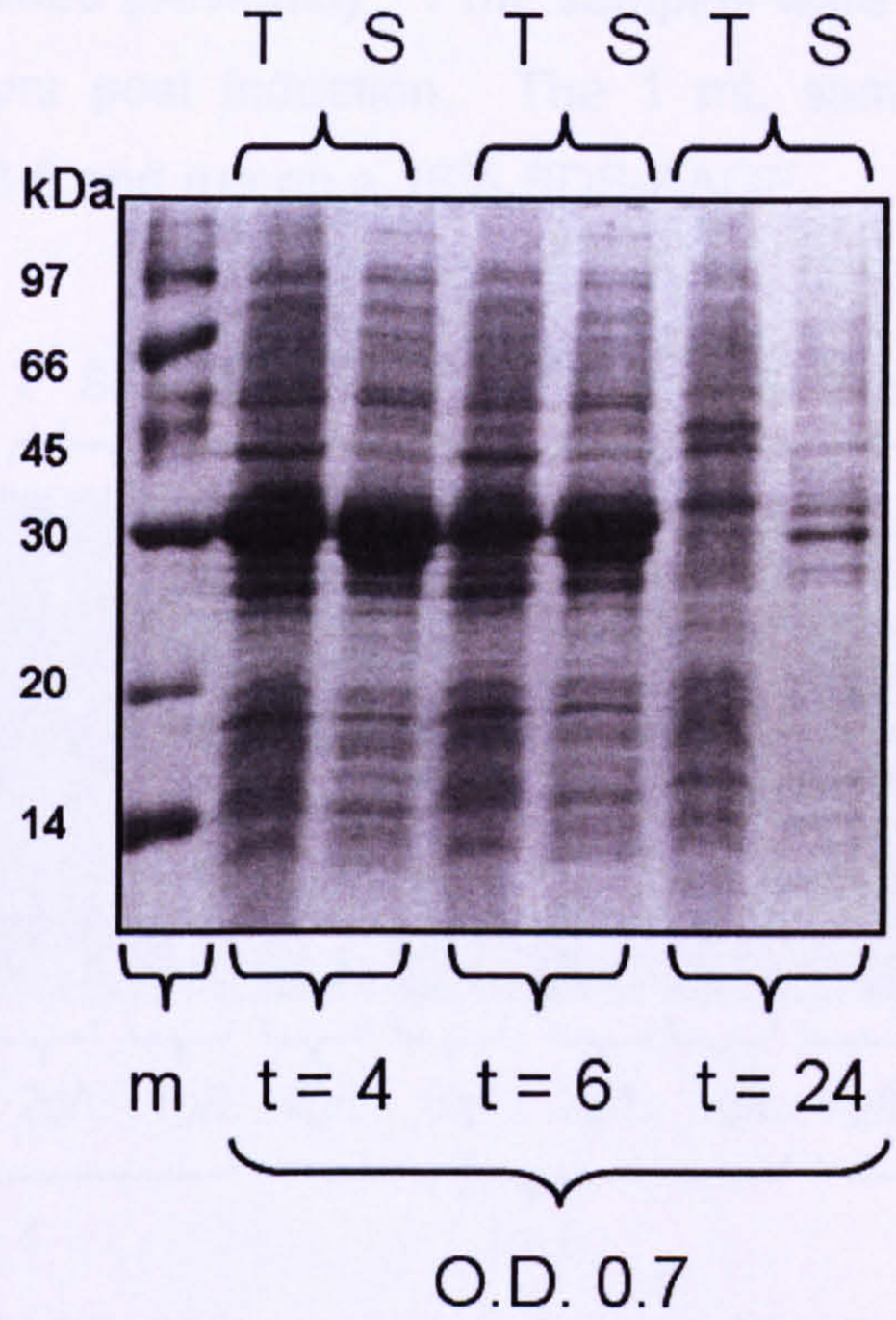

Figure 3-27. 16% SDS-PAGE of bb'x test expression. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), O.D. = optical density measured at 600nm at which cells were induced, m = low molecular weight marker (kDa).

From Figure 3-26 and Figure 3-27 it was clear that induction at 0.3 causes a huge reduction in protein expression when compared to lanes with induction at optical densities of 0.5 and 0.7. When comparing induction at optical densities of 0.5 and 0.7, the expression was perhaps slightly higher at O.D. 0.5. When comparing expression time at O.D. 0.5, it is very clear that the level of expressed protein is at its highest concentration at 4 hours and there is a marked decrease in protein after 6 hours of expression and almost no protein after 24 hours of expression.

To measure the effect of glucose concentration on yield, four 50 mL minimal medium cultures were grown on differing concentrations of glucose of 4, 3, 2 and 1 g/L and induced with 1 mM IPTG at O.D. of 0.5, highlighted by Figure 3-26 and Figure 3-27 as being the optimum O.D. for induction for maximum

protein yield; determined previously. 1 mL samples were taken for each culture at 4, 6 and 24 hours post induction. The 1 mL samples were treated as described in Figure 3-6 and run on a 16% SDS-PAGE.

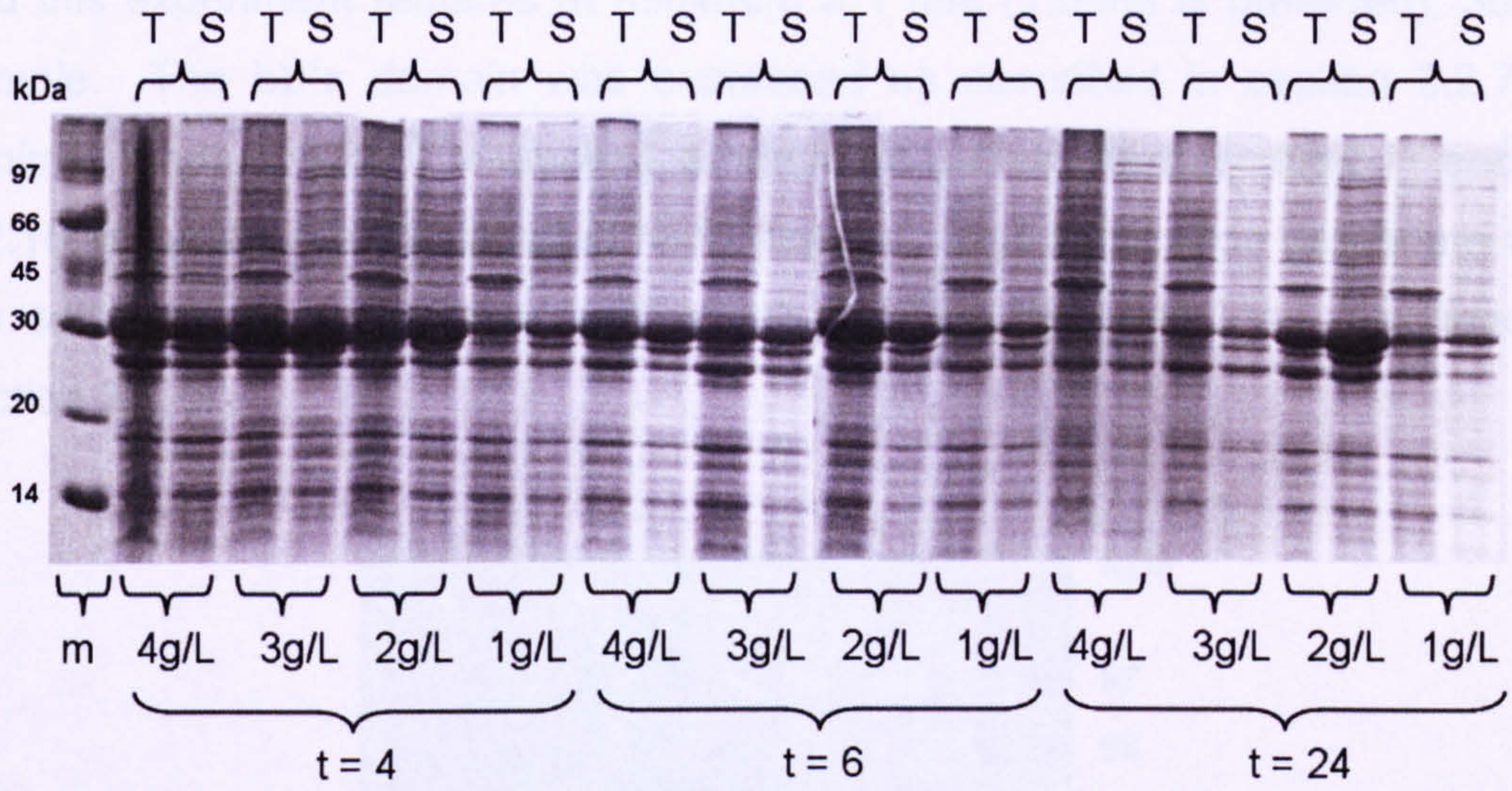

Figure 3-28. 16% SDS-PAGE of bb'x glucose test expression. T = Total protein sample, S = Soluble protein sample, Glucose concentration in g/L (grams per Litre), t = time of sample taken after induction (Hours), m = low molecular weight marker (kDa).

From Figure 3-28 it is clear that 24 hour induction results in a much lower yield of protein expressed. The samples taken at 2g/L at 24 hours appear anomalous in that the level of protein expression appears much higher than another sample taken over a 24 hour induction period. When comparing 4 hour and 6 hour inductions it is also clear that the expressed **bb'x** protein is at its highest concentration in the 4 hour samples. Comparing the soluble protein samples during the 4 hours inductions, it is clear that 1 g/L expresses the least amount of protein. Surprisingly, 4 g/L appears to expresses less protein than 3 and 2g/L, where there is less glucose present (this could be an anomalous result, resulting from poor handling handling of this sample). It was difficult to distinguish any great difference between the size of the bands in soluble protein samples of 3 and 2g/L, but the sample at 3 g/L appears slightly larger. In the interest of cost effectiveness, the slight drop in expression in using 2 g/L was acceptable, since the difference in the expression level appeared slight.

3.3.3. Expression and purification of ^{15}N bb'x

To first assess the quality of NMR data, a 2D HSQC experiment was necessary and this experiment requires at minimum a 1 mM (1.5mM is preferred), 300uL sample. The **bb'x** domain was expressed as described in section 2.2.7, in minimal medium with ^{15}N-labelled ammonium sulfate as described in section 2.2.10, to produce a ^{15}N-labelled NMR sample. The purification procedure was the same as all expression purifications, starting with IMAC as described in section 2.3.1.

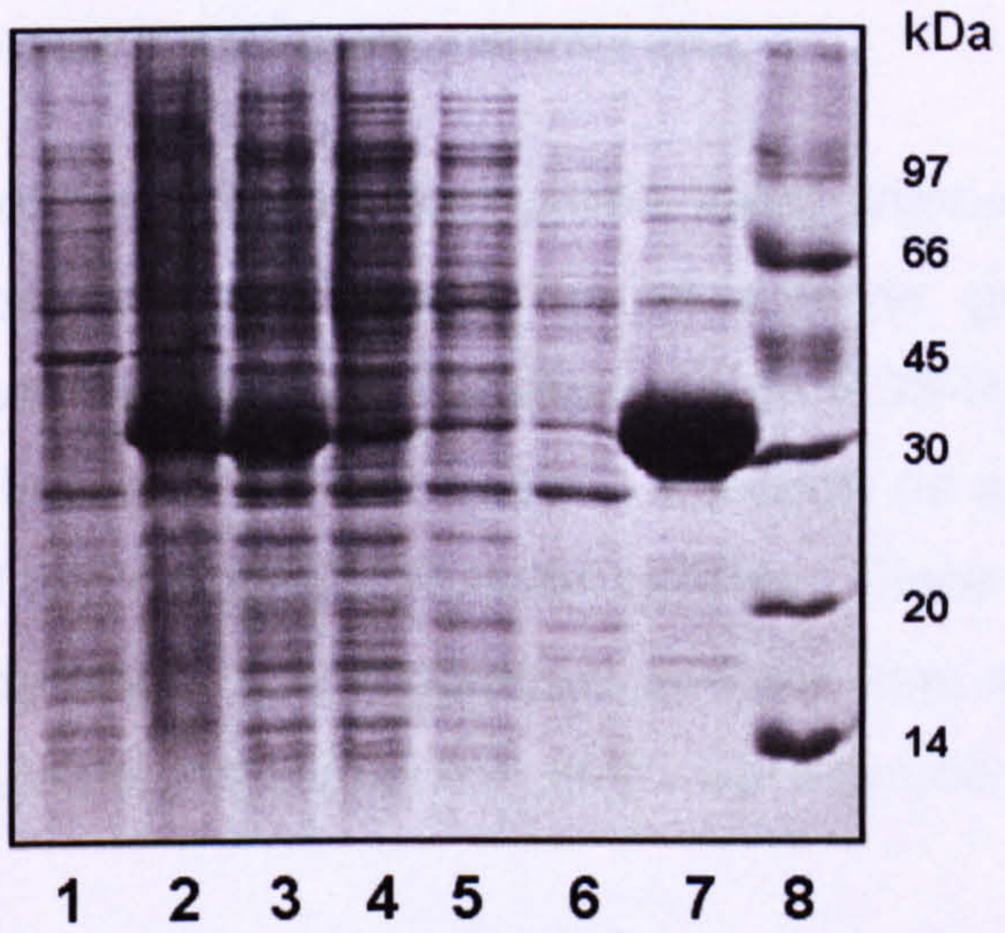

Figure 3-29. 16 % SDS-PAGE of ^{15}N bb'x applied to IMAC. Lane 1, un-induced sample. Lane 2, total protein. Lane 3, soluble protein. Lane 4, flow through. Lane 5, imidazole wash. Lane 6, low salt wash. Lane 7, hexa-histidine tagged bb'x protein elute. Lane 8, low molecular weight marker.

The IMAC purification step was very effective at removing the majority of protein impurities. But the elute still has several protein impurities remaining. The IMAC elute was then further purified using anion-exchange chromatography as described in section 2.3.2.

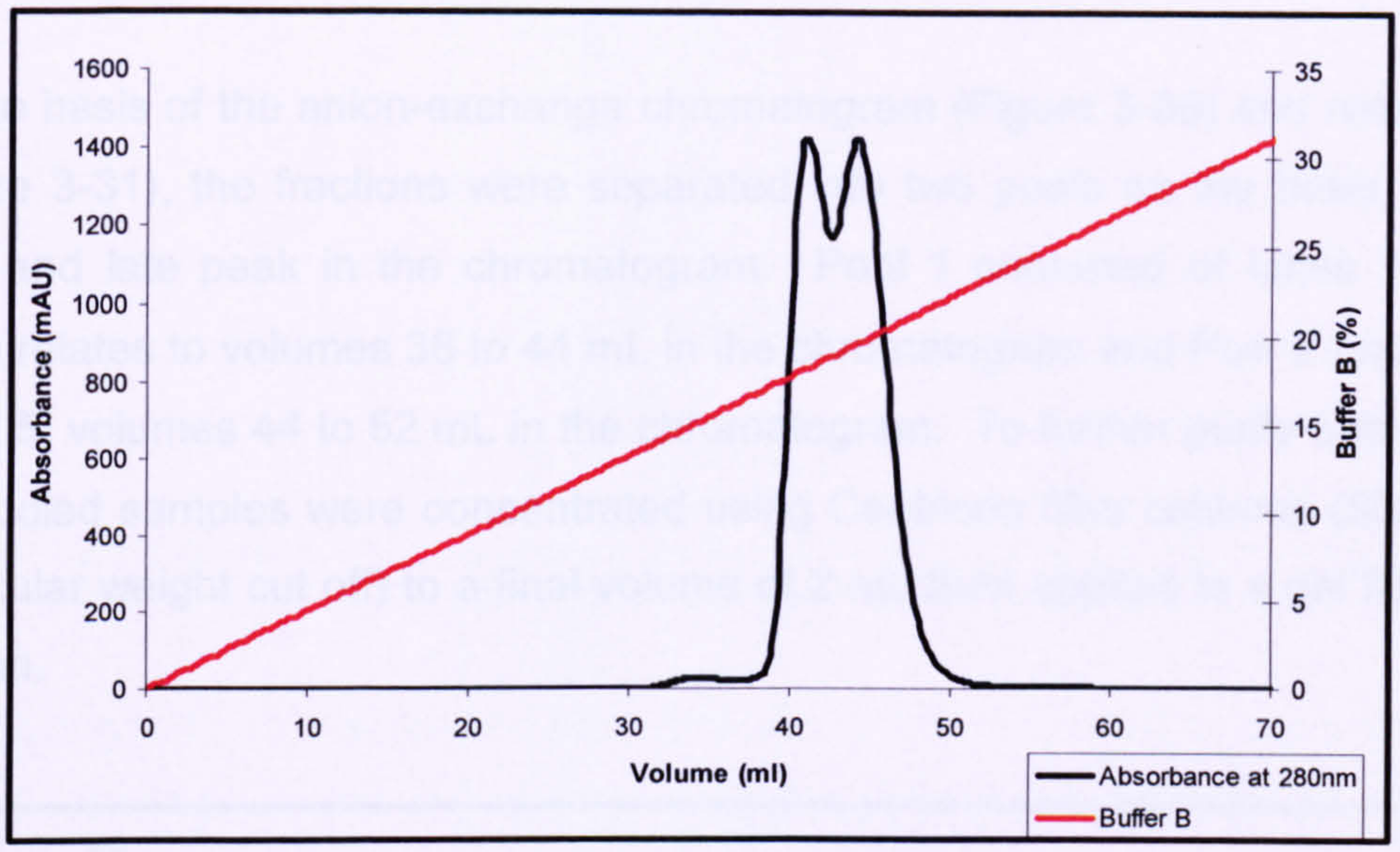

Figure 3-30. Chromatogram of bb'x using a Source 30Q.

The chromatogram of the ^{15}N-**bb'x** purification using anion-exchange in Figure 3-30 reveals two peaks even though run at a lower gradient to improve resolution. But as seen in Figure 3-15 when run on SDS-PAGE there appears to be no difference between the two peaks. Although as seen previously, the two peaks represent two differing molecular species. Since this sample was to be used analysed by NMR, a homogenous species was essential and so to determine this, 10 µL samples across the peaks were run on native-PAGE.

Figure 3-31. 16 % native-PAGE of bb'x from fractions of the peak in Figure 3-30, from 38 mL to 48 mL eluted volume.

On the basis of the anion-exchange chromatogram (Figure 3-30) and native gel (Figure 3-31), the fractions were separated into two pools on the basis of the early and late peak in the chromatogram. Pool 1 consisted of lanes 1 to 3, which relates to volumes 38 to 44 mL in the chromatogram and Pool 2 relates to 4 and 5, volumes 44 to 52 mL in the chromatogram. To further purify each pool, the pooled samples were concentrated using Centricon filter columns (5000 Da molecular weight cut off) to a final volume of 2 mL then applied to a gel filtration column.

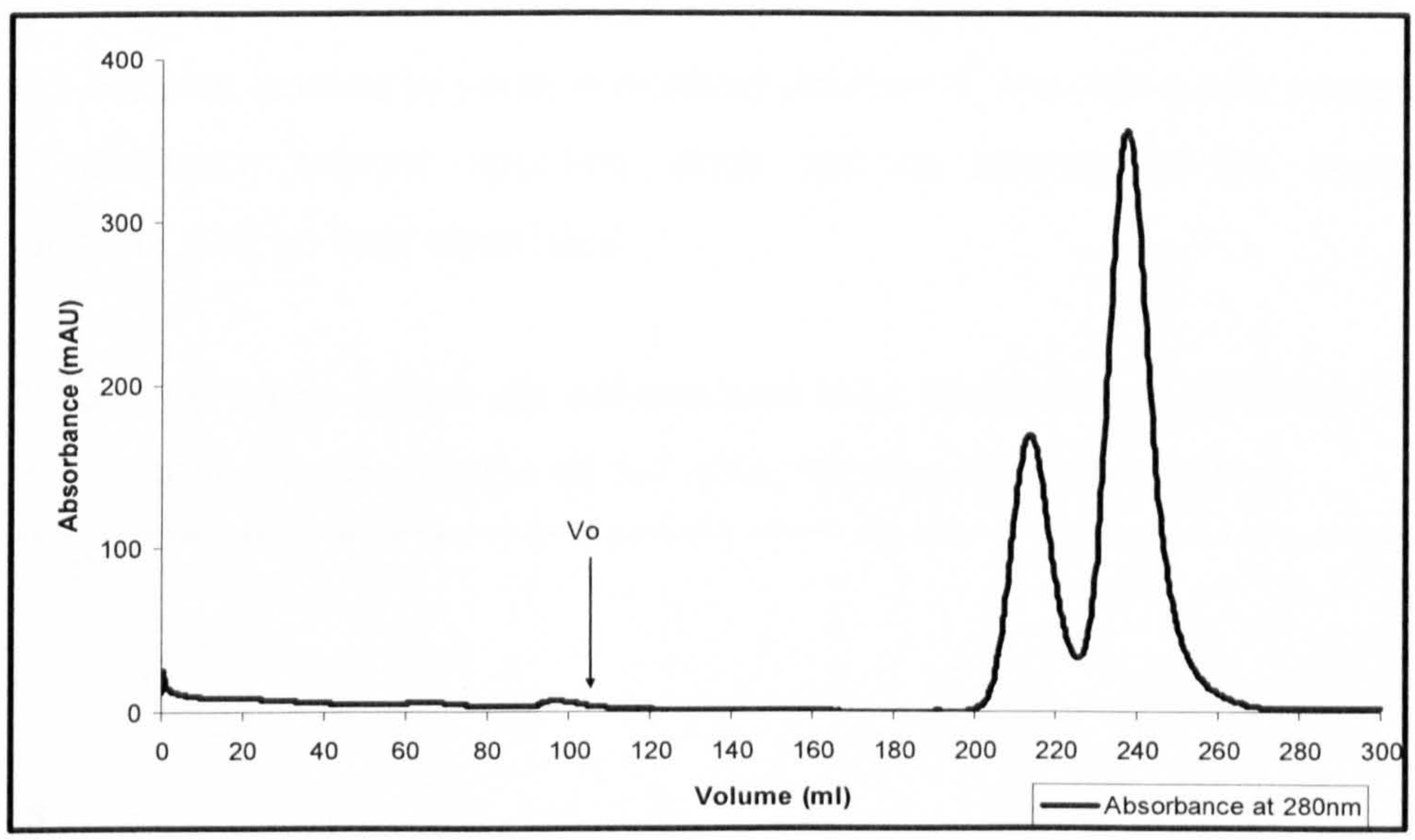

Figure 3-32. Superdex 75 gel filtration chromatogram of Pool 1. The void volume highlighted with V_o is 105mL

From Figure 3-32 it is clear that the initial anion-exchange Pool 1 separation was not very effective at separating the two species into a single homogenous sample. The separation by gel filtration appears to be more effective at separating the monomer and dimer. To confirm the homogeneity of the two peaks, 20 µL samples were taken from each peak and analysed by native-PAGE.

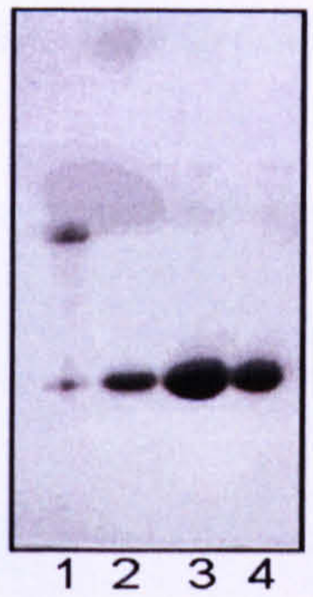

Figure 3-33. 16 % native-PAGE of bb'x from fractions of the peaks in Figure 3-32. Lane 1, fraction containing first peak. Lane 2 to 4 fractions across second peak.

The homogeneity of the second peak (monomer) was revealed by the native gel. So the second peak (lower molecular weight species) from the chromatogram in Figure 3-32 was pooled in what is defined as Pool 1, the first peak sample (the higher molecular weight species) does not to appear to be completely homogenous and so was discarded.

Pool 2 from the anion-exchange column was then applied to gel filtration column, again to improve the separation of the initial anion-exchange pooling.

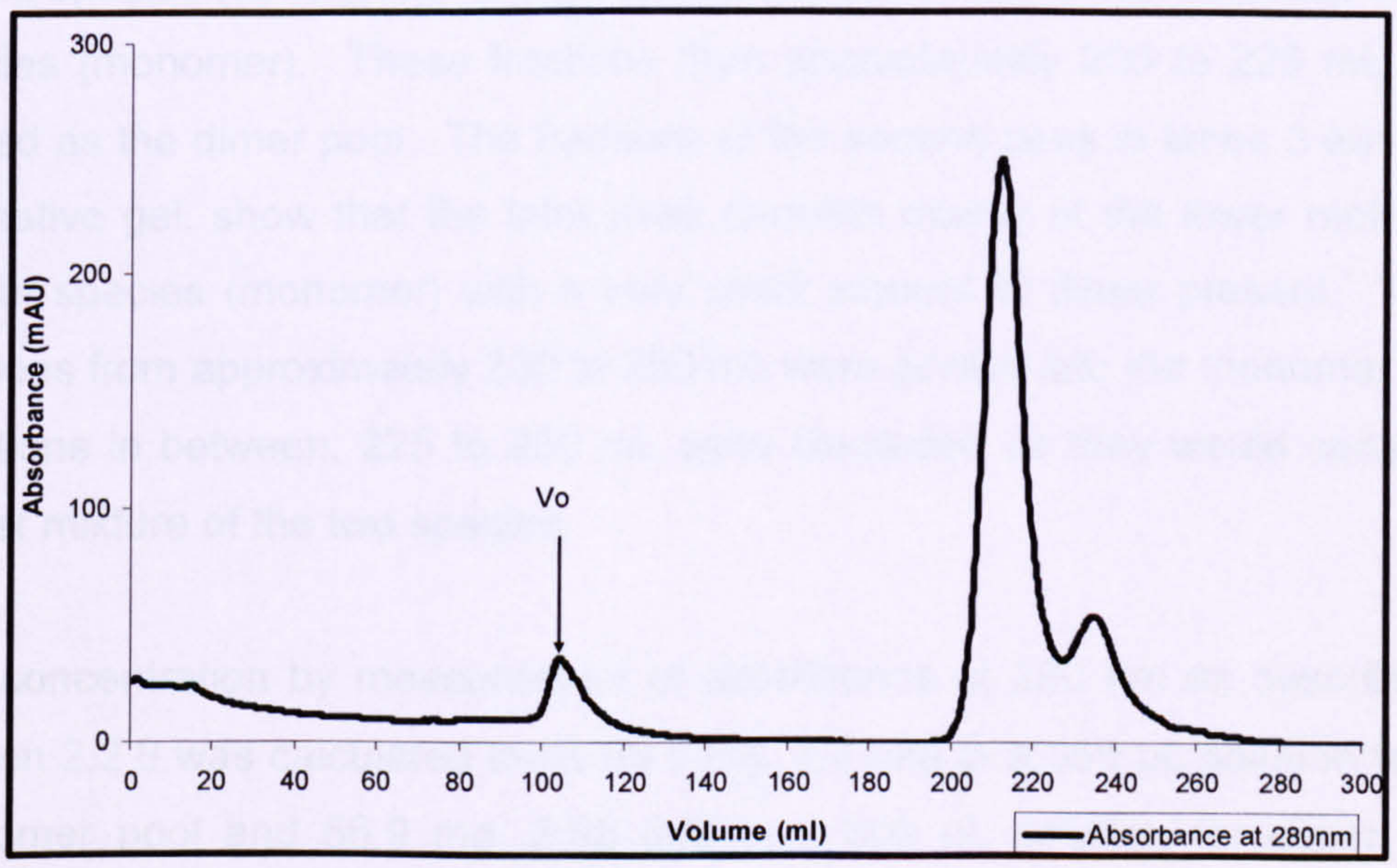

Figure 3-34. Superdex 75 gel filtration chromatogram of Pool 2. The void volume highlighted with V_o is at 105mL.

To check the identity of each peak, again samples from each peak were taken and run on a native gel.

Figure 3-35. 16 % native-PAGE of bb'x from fractions of the peaks in Figure 3-34. Lane 1, fraction containing first peak. Lane 2, empty. Lane 3 and 4, fractions across second peak.

As shown in the native gel, the first peak contains mainly the larger molecular weight species (dimer), with a slight contamination of the lower molecular weight species (monomer). These fractions from approximately 200 to 225 mL were pooled as the dimer pool. The fractions of the second peak in lanes 3 and 4 on the native gel, show that the later peak consists mainly of the lower molecular weight species (monomer) with a very small amount of dimer present. These fractions from approximately 230 to 250 mL were pooled into the monomer pool. Fractions in between; 225 to 230 mL were discarded as they would consist of higher mixture of the two species.

The concentration by measurement of absorbance at 280 nm as described in section 2.2.8 was calculated to be 63.9 mg, 2.3 mM in a 350 µL solution for the monomer pool and 56.9 mg, 2.05 mM in a 500 µL solution for dimer pool. Samples of each were then prepared and analysed by mass spectrometry.

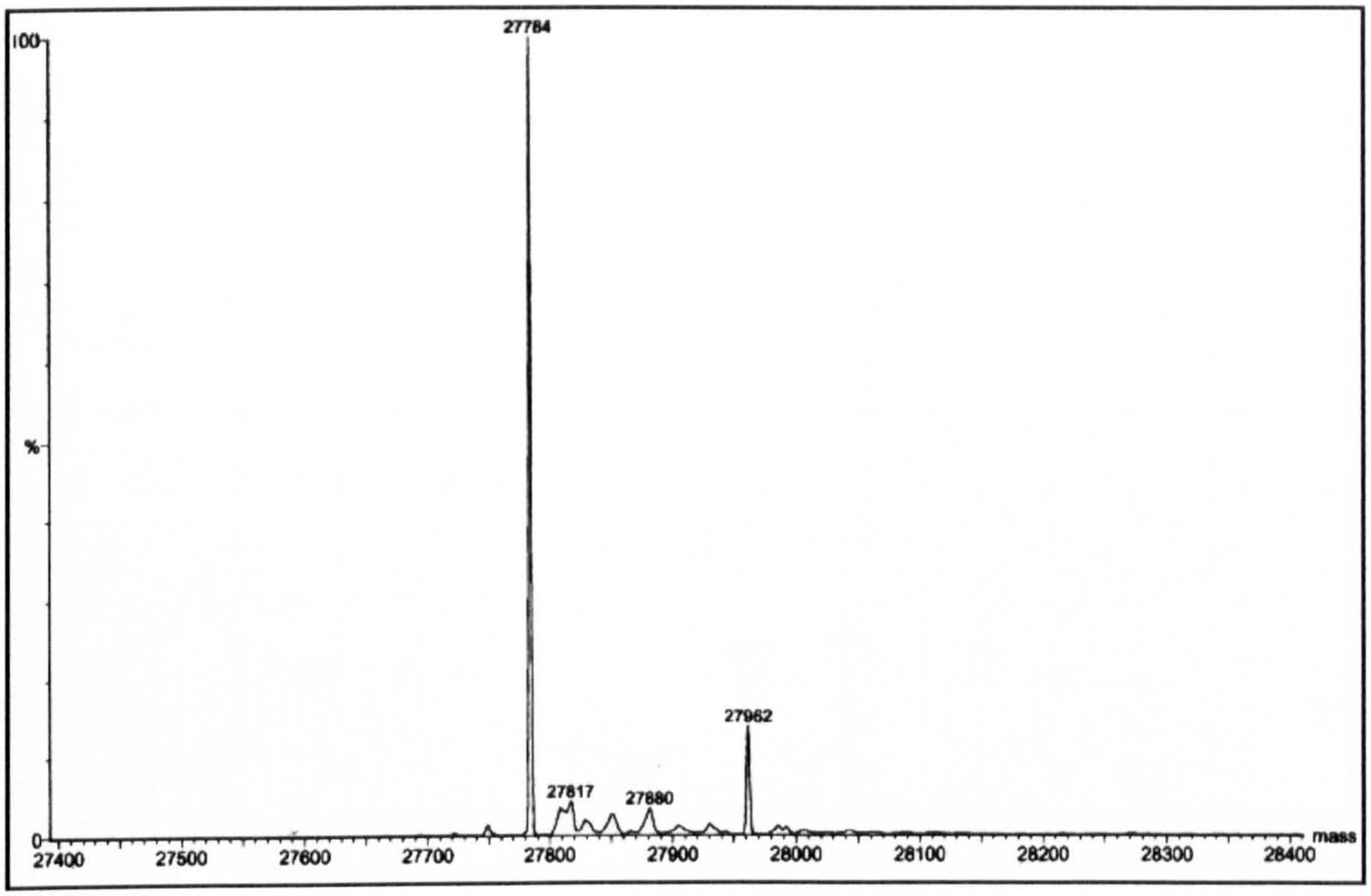

Figure 3-36. Positive ion mass spectrum of ^{15}N bb'x monomer pool.

The expected molecular weight if all 316 Nitrogen atoms were labelled with the ^{15}N isotope would be 27793 Da since the unlabelled **bb'x** domains is 27477 Da. The actual molecular weight recorded electrospray mass spectrometry was 27784.0 Da, which indicates 307 labelled Nitrogen of the 316; therefore the labelling efficiency was 97% in the monomer sample. The mass spectrum in Figure 3-36 also shows some minor species as discussed previously.

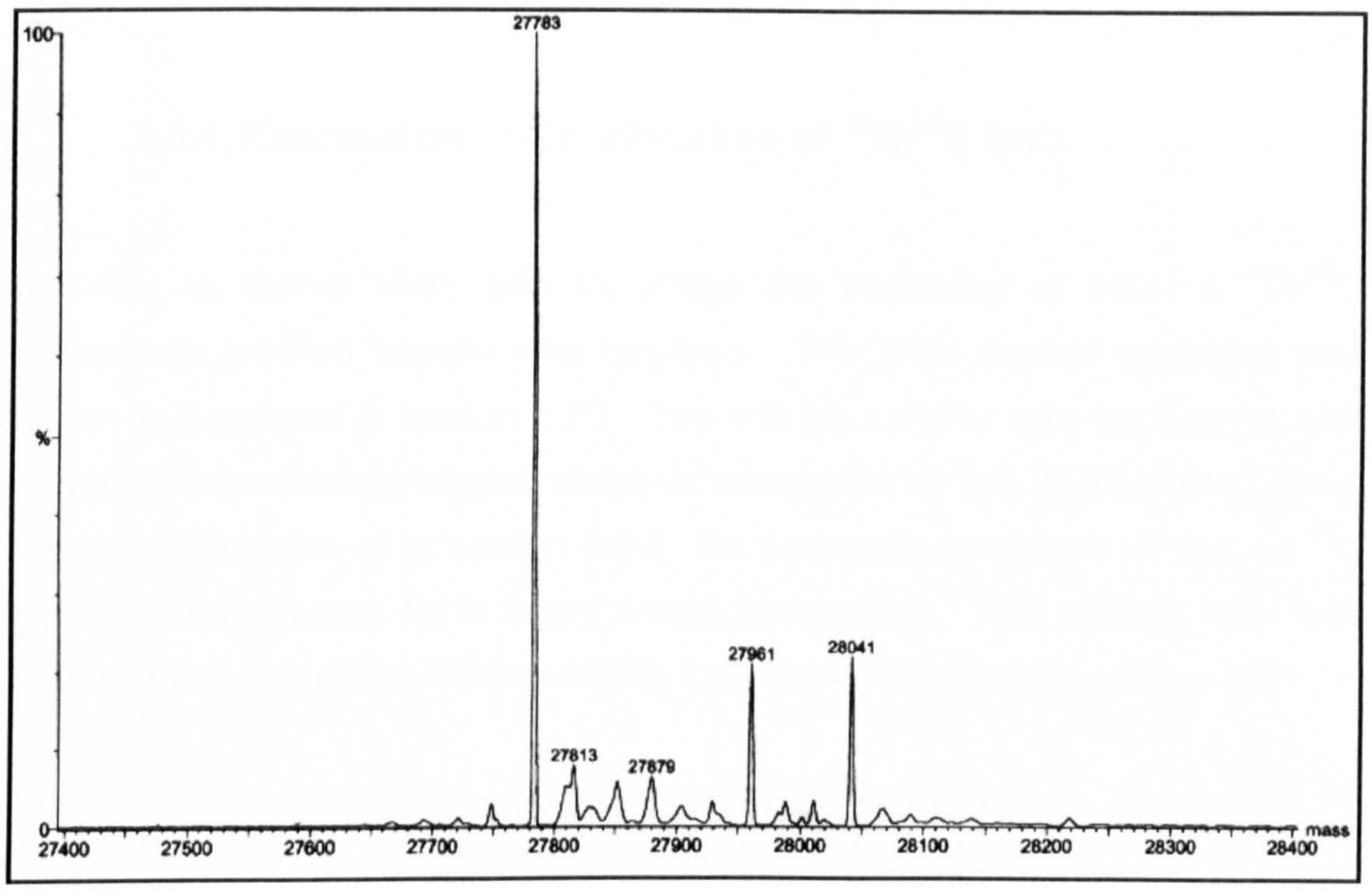

Figure 3-37. Positive ion mass spectrum of ^{15}N bb'x dimer pool.

100% labelling of all the 316 Nitrogen atoms would reveal a species at 27793 Da since the unlabelled **bb'x** domains is 27477 Daltons. The actual molecular weight recorded by electrospray mass spectrometry was 27783 Da, which indicates 306 labelled Nitrogen of the 316; therefore the labelling efficiency was 96.8% in the dimer pool sample.

The mass spectrum in Figure 3-37 also reveals the dephosphorylated (+178 Da) and phosphorylated (+258 Da) α-N-6-Phosphogluconoylation of the histidine-tag modification at 27961 Da and 28041 Da respectively. There was also some of the other minor species as discussed previously. The phosphorylated (+258 Daltons) α-N-6-Phosphogluconoylation of the histidine tag has so far only appeared in the dimer samples, seen here in Figure 3-37 and seen in the unlabelled **bb'x** domain sample in Figure 3-24, the importance or relevance of this observation was unclear and could just be a coincidental.

3.3.4. Expression and purification of $^{15}N/^{13}C$ bb'x

In order to collect NMR data to assign the backbone of **bb'x**, a $^{15}N/^{13}C$ isotopically labelled sample was required. The **bb'x** domain construct was grown as described in section 2.2.7. Two 400 mL minimal medium cultures with appropriate isotopically labelled reagents were grown in two 2 Litre flasks. From the test expression of in section 3.3.2, the optimised conditions of 2g/L of ^{13}C glucose and induced for 4 hours would be applied. The sample was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1.

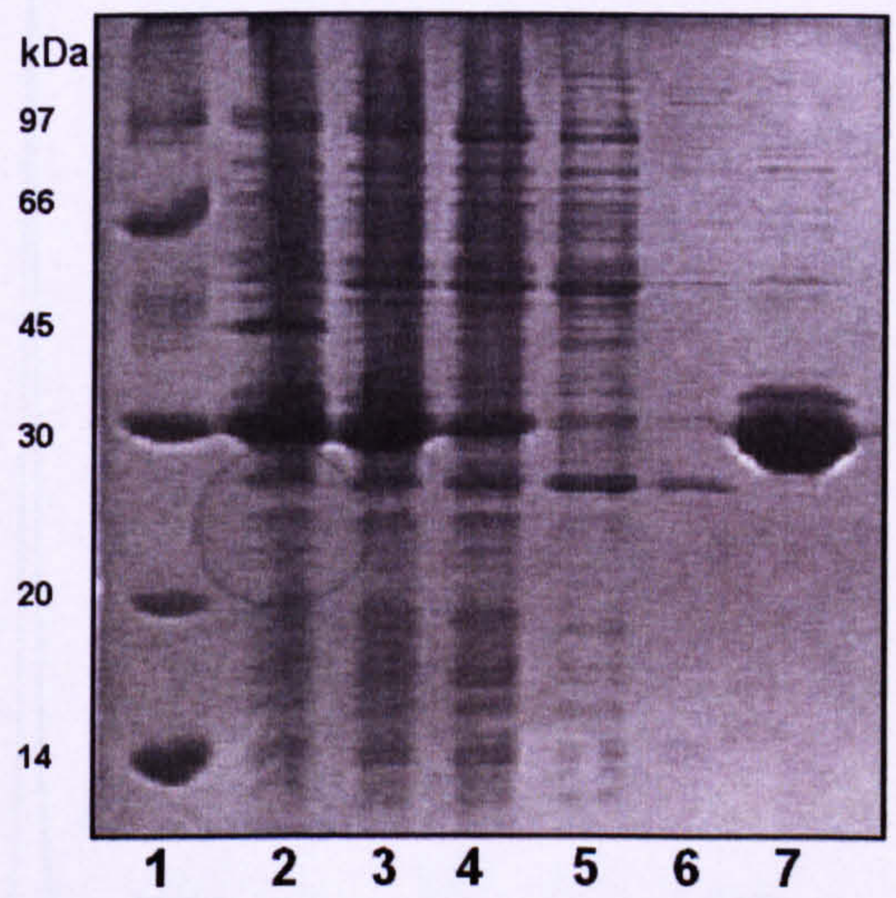

Figure 3-38. 16% SDS-PAGE of the IMAC purification of $^{15}N/^{13}C$ bb'x. Lane 1, low molecular weight marker. Lane 2, total protein. Lane 3, soluble protein. Lane 4, flow through. Lane 5, imidazole wash. Lane 6, low salt wash. Lane 7, hexa-histidine tagged $^{15}N/^{13}C$ bb'x protein elute

From Figure 3-38, it is clear that purification step was very effective at removing the majority of non-tagged protein impurities. But the elute still has several minor protein impurities remaining. This $^{13}C/^{15}N$ **bb'x** sample was further purified by anion-exchange chromatography, this $^{13}C/^{15}N$ **bb'x** gave comparative data to the ^{15}N **bb'x** sample as described in section 3.3.3.

The $^{13}C/^{15}N$ **bb'x** sample was then separated into two pools on the basis of the gel filtration peaks and native-PAGE analysis, similar to as was done in section 3.3.3. The two pools were then concentrated down and buffer exchange into the NMR buffer. The concentration by measurement of absorbance at 280nm and was calculated to be 179.1 mg, 6.17 mM in a 200 µL sample for monomer pool. For the dimer pool the concentration was calculated, using the same method, 29.0 mg, 1.0 mM in a 200 µL sample. A small sample of monomer and dimer was buffer exchanged into 10 mM ammonium bicarbonate and analysed by mass spectrometry.

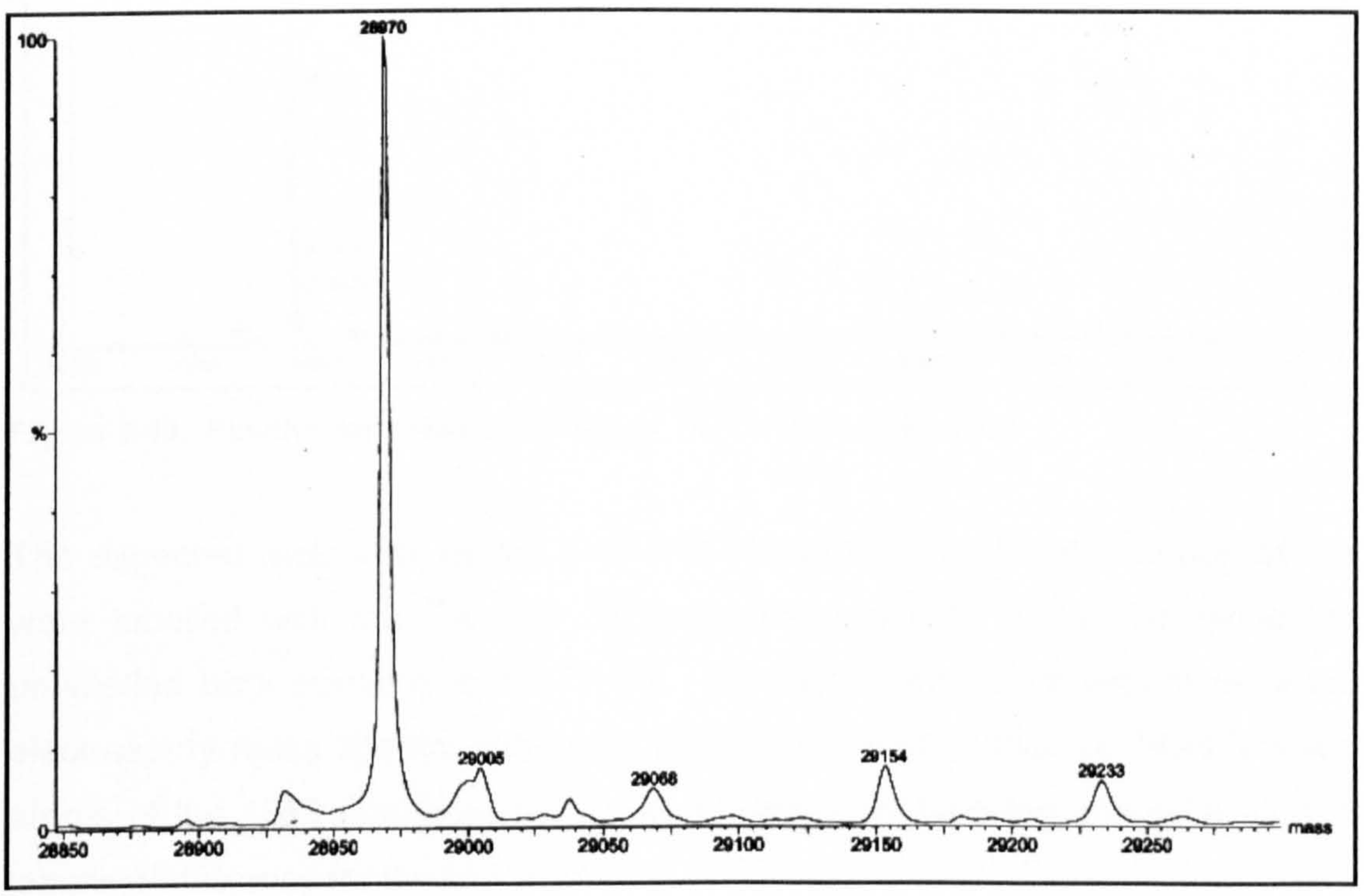

Figure 3-39. Positive ion mass spectrum of $^{13}C/^{15}N$ bb'x monomer pool.

The expected molecular weight if all the 316 Nitrogen and all the 1242 Carbon atoms were labelled with the ^{15}N and ^{13}C isotopes would be 29035 Da since the unlabelled **bb'x** domains is 27477 Da. The actual molecular weight recorded electrospray mass spectrometry was 28970, Da which indicates 1493 labelled

atoms of the 1558 total possible atoms available for labelling and so a 95.8 % labelling efficiency for the monomer pool. Several other minor species as described previously are also recorded.

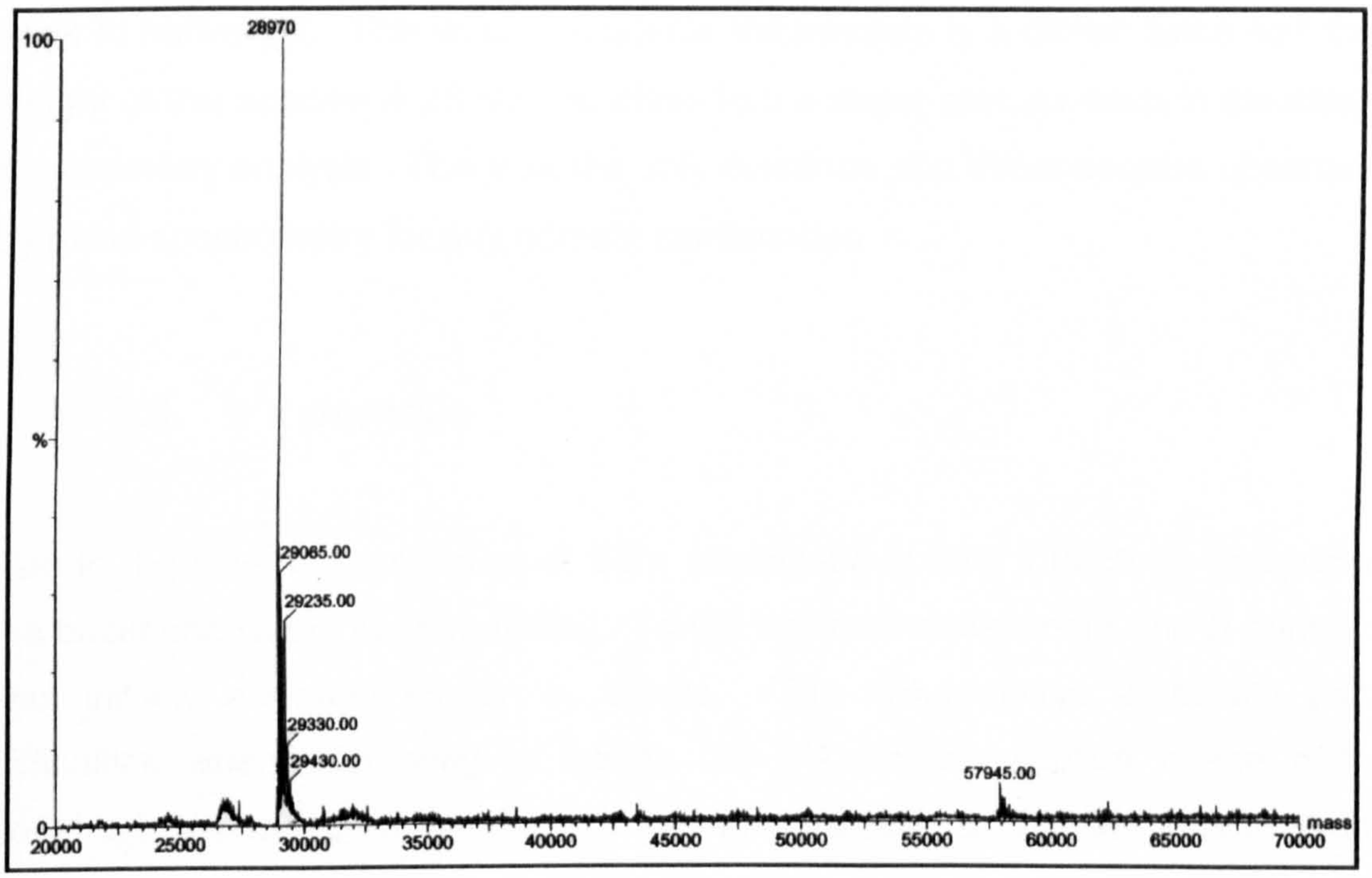

Figure 3-40. Positive ion mass spectrum of $^{13}C/^{15}N$ bb'x dimer pool.

The expected molecular weight if all 316 Nitrogen and all 1242 Carbon atoms were labelled with the ^{15}N and ^{13}C isotopes would be 29035 Da since the unlabelled **bb'x** domains is 27477 Da. The actual molecular weight recorded electrospray mass spectrometry was 28970 Da, which indicates 1493 labelled atoms of the 1558 total possible atoms available for labelling and so a 95.8 % labelling efficiency for the dimer pool.

There are also several other species which appear in the mass spectrometry analysis, the peak at 29065 Da results from a possible phosphorylated adduct with an extra mass of 95 Da. The species at 29235 Da, with an extra mass of 265 Da could result from the α-N-6-Phosphogluconoylation of the histidine tag modification (258 Da), the extra 7 Dalton mass difference could result from six ^{13}C or ^{15}N atoms.

Most interestingly and small amount of a species with a mass of 57945 was observed. This could result from the high molecular weight species or dimer seen in native gel. This weight suggests the species is a dimer, since half the weight of this species is 28973 Da, close to the major species seen in the mass spectrometry analysis. This was the only evidence of a dimer species observed by mass spectrometry for any domain combination.

3.4. b'x domain

Due to the size and properties of **bb'x** (discussed in later Chapters) assigning the backbone would be demanding. To aid accurate assignment, the **b** domain was initially assigned by Dr. K. Wallis. This would leave ambiguity and difficulties when attempting to assign the **b'x** domain section in the **bb'x** construct especially since the **b'** domain appears to be in conformation exchange; possibly resulting from functional requirements. Therefore, it would be beneficial to also initially assign **b'x** before tackling **bb'x**.

3.4.1. Expression and purification of ^{15}N b'x

The optimisation and test expression was carried out in the group by Dr. K. Wallis and was found to be the same as those described in section 2.2.7. Also carried out was mass spectrometry analysis, which confirmed the 17282 Da mass of the **b'x** domain construct identical to the predicted mass from the sequence.

In order to isotopically label **b'x** the ammonium sulphate in the minimal medium was substituted with ^{15}N-labelled ammonium sulphate as described in section 2.2.10. Two 400mL cultures were grown in two 2 Litre flasks. The clarified cell

extract was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1.

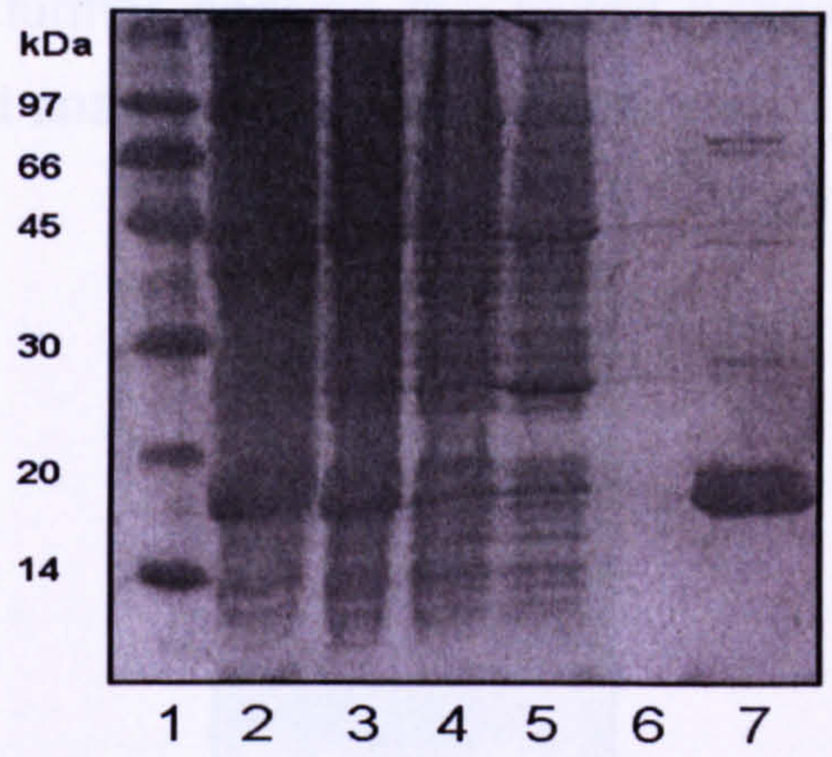

Figure 3-41. 16% SDS-PAGE of the IMAC purification of ^{15}N b'x. Lane 1, low molecular weight marker. Lane 2, total protein. Lane 3, soluble protein. Lane 4, flow through. Lane 5, imidazole wash. Lane 6, low salt wash. Lane 7, hexa-histidine tagged ^{15}N b'x protein elute.

From Figure 3-41, it was clear that purification step was very effective at removing the majority of non-tagged protein impurities. But the elute still has several minor protein impurities remaining. The elute sample was then further purified using anion-exchange chromatography as described in section 2.3.2.

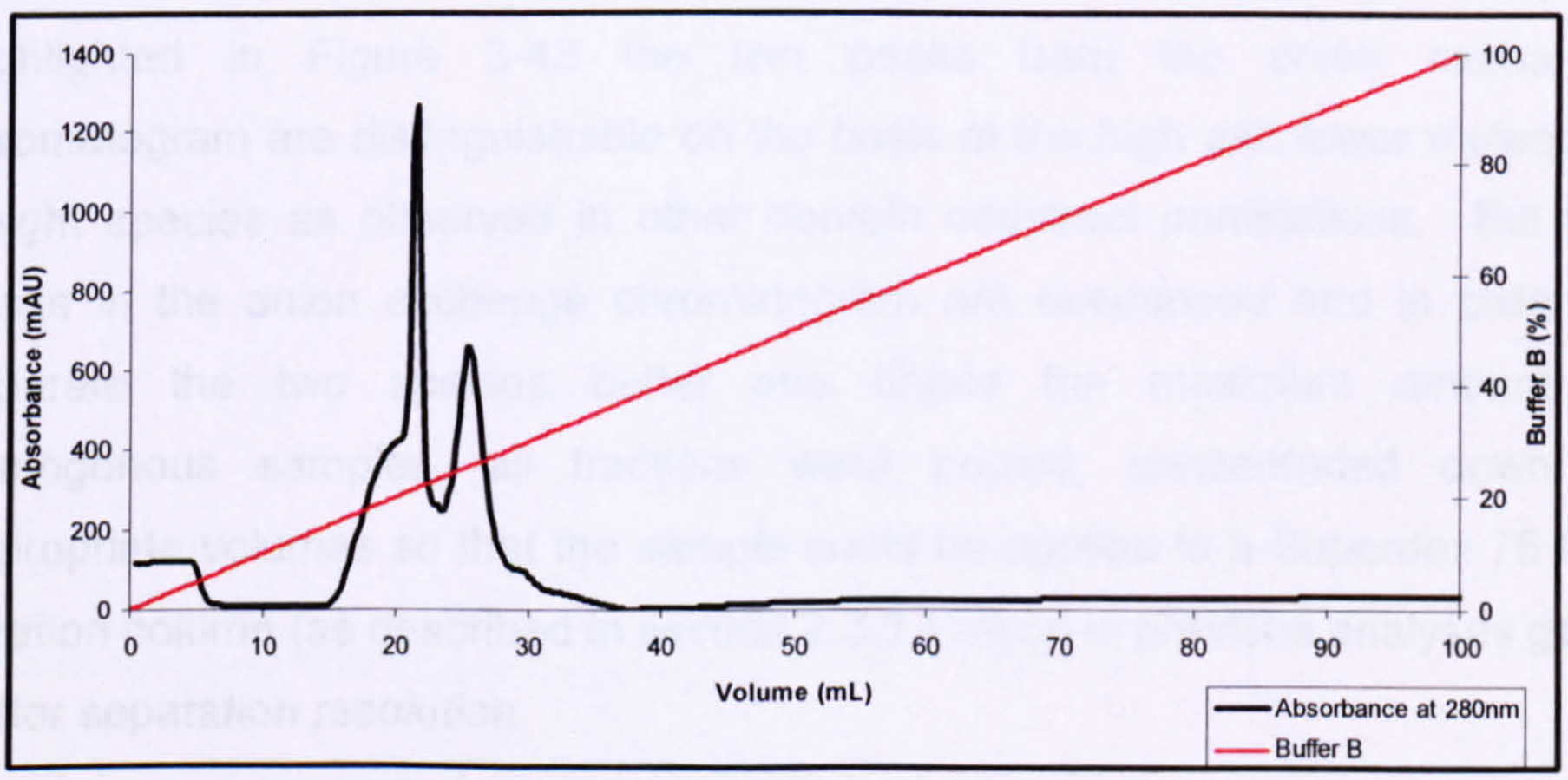

Figure 3-42. Chromatogram of ^{15}N b'x using a Source 30Q with a 100% Buffer B gradient over 100mL.

The chromatogram of the anion-exchange reveals two major peaks; also this chromatogram was somewhat different from all other PDI domain purification chromatograms. To further analyse the eluted protein, samples were taken across both peaks and analysed by native-PAGE.

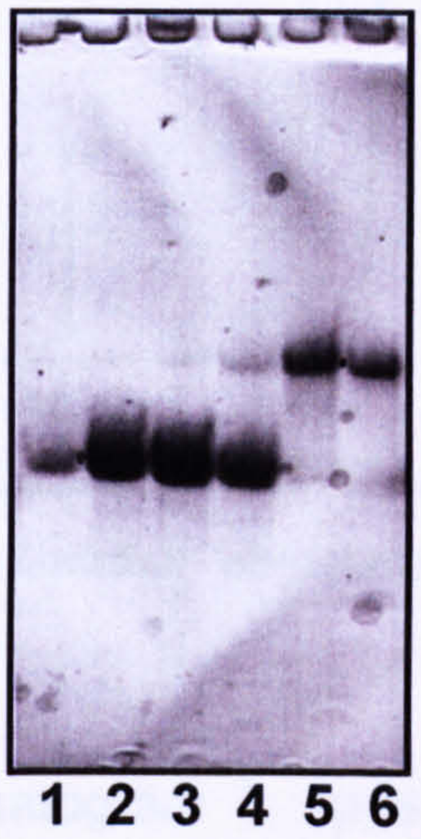

Figure 3-43. 16% native-PAGE of fractions across peaks in Figure 3-42. Lanes 1- 3 taken from across the early peak, Lane 4 taken from a fraction between the two peaks and Lanes 5,6 from fraction of the late peak.

Since this protein was an expensive labelled preparation, only small number and volume of selected samples were used for native-PAGE analysis. Highlighted in Figure 3-43 the two peaks from the anion exchange chromatogram are distinguishable on the basis of the high and lower molecular weight species as observed in other domain construct purifications. But the peaks in the anion exchange chromatogram are overlapped and in order to separate the two species better and obtain the maximum amount of homogenous samples, all fractions were pooled, concentrated down to appropriate volumes so that the sample could be applied to a Superdex 75 Gel filtration column (as described in section 2.3.3.) which in previous analyses gave better separation resolution.

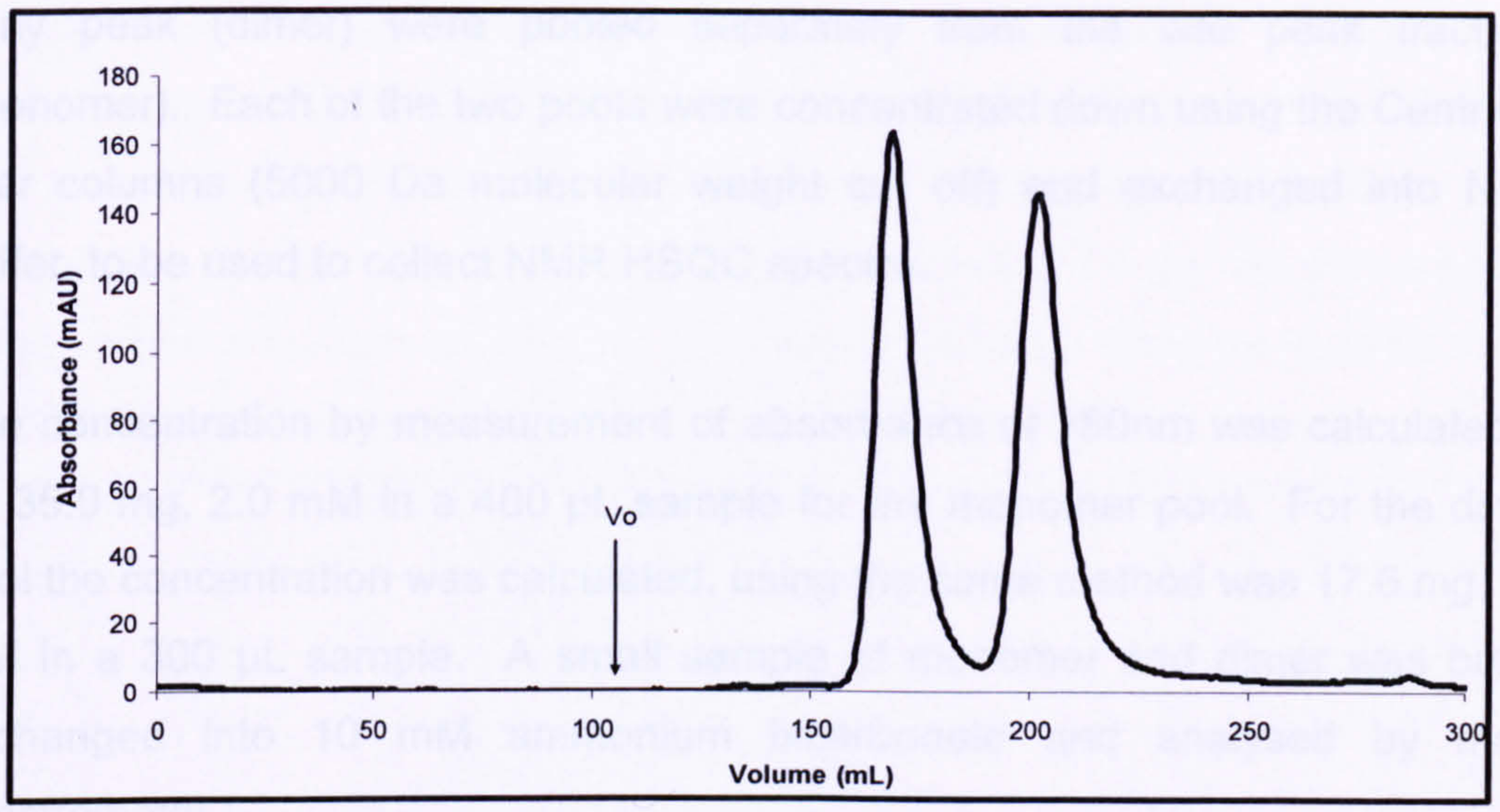

Figure 3-44. Superdex 75 gel filtration chromatogram ^{15}N b'x. The void volume highlighted with V_o at 105mL.

From the gel filtration chromatogram it appears the two forms have been separated into two homogenous samples. To ensure this was the case samples from each of the two peaks were analysed by Native-PAGE.

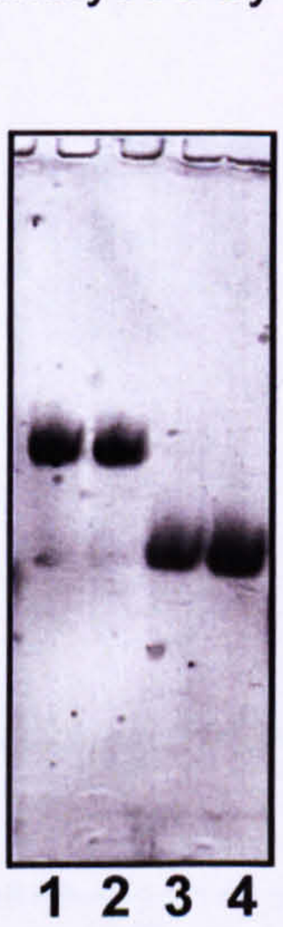

Figure 3-45. 16% native-PAGE of fractions across each peak in Figure 3-44. Lanes 1-2 from the early gel-filtration peak and Lanes 3-4 from the late peak.

From the native-PAGE analysis taken from samples from each peak separated by gel-filtration it was clear that monomer and dimer species are well separated and result in homogenous fractions. Hence all fractions from the gel-filtration

early peak (dimer) were pooled separately from the late peak fractions (monomer). Each of the two pools were concentrated down using the Centricon filter columns (5000 Da molecular weight cut off) and exchanged into NMR buffer, to be used to collect NMR HSQC spectra.

The concentration by measurement of absorbance at 280nm was calculated to be 35.0 mg, 2.0 mM in a 400 µL sample for the monomer pool. For the dimer pool the concentration was calculated, using the same method was 17.6 mg, 1.0 mM in a 300 µL sample. A small sample of monomer and dimer was buffer exchanged into 10 mM ammonium bicarbonate and analysed by mass spectrometry.

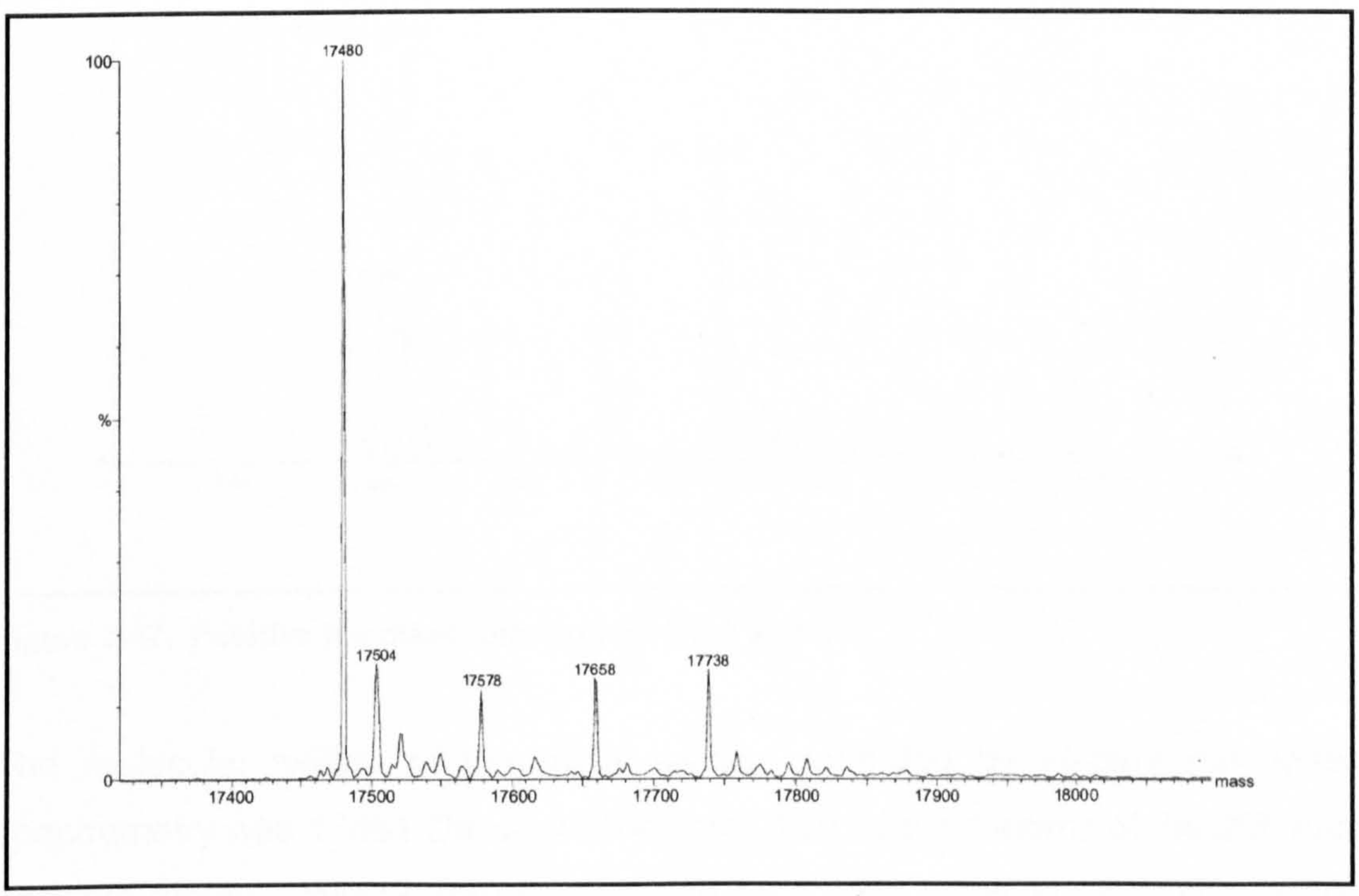

Figure 3-46. Positive ion mass spectrum of ^{15}N b'x monomer.

The expected molecular weight if all 205 Nitrogen were labelled with the ^{15}N-isotopes would be 17487 Da since the unlabelled **b'x** domains is 17282 Da. The molecular weight of the main species recorded by electrospray mass spectrometry was 17480 Da, which indicates 198 labelled atoms of the 205 total

possible atoms available for labelling and so a 97% labelling efficiency for the monomer pool. The 17658 Da and 17738 Da species have an extra mass of 178 and 256 Da respectively, resulting from the His-tag modification. The species with a mass of 17504 Da has an extra mass of 24 Da and so was likely to result from a sodium ion adduct. The species with an extra mass of 98 Da at 17578 Da, was likely to result from a phosphorylation.

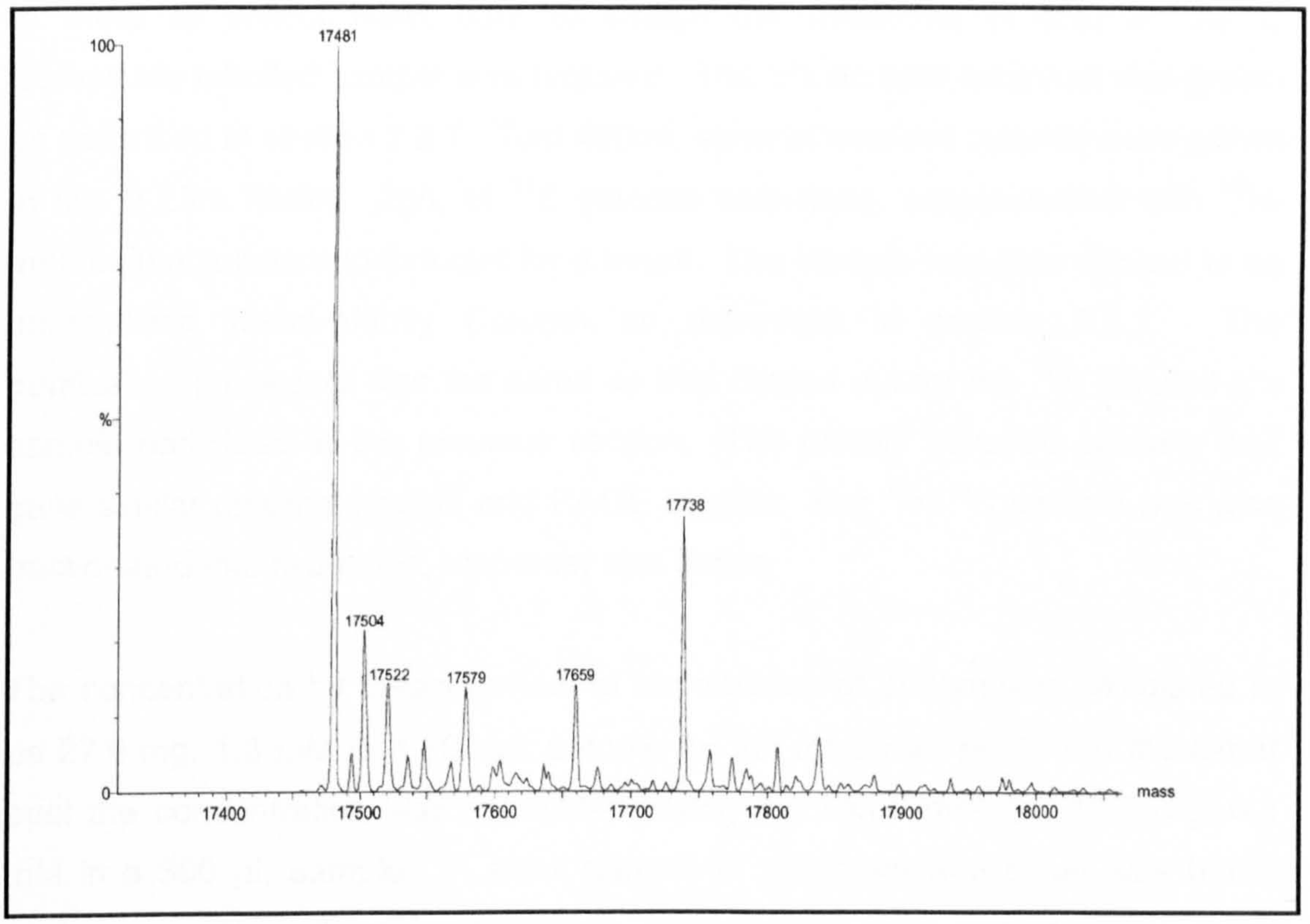

Figure 3-47. Positive ion mass spectrum of ^{15}N b'x dimer.

The molecular weight of the main species recorded by electrospray mass spectrometry was 17481 Da, which indicates 199 labelled atoms of the 205 total possible atoms available for labelling and so a 97% labelling efficiency for the dimer pool. The 17659 Da and 17738 Da species have an extra mass of 178 Da and 257 Da respectively, resulting from the His-tag modification. The species with a mass of 17504 Da has an extra mass of 24 Da, likely to result from a single sodium ion adduct and the mass at 17522 Da (extra 41 Da) likely to result from two sodium ions. The species with an extra mass of 98 Da at

17579 Da, was likely to result from a phosphorylation as seen in previous mass spectrometry analysis.

3.4.2. Expression and purification of $^{13}C/^{15}N$ b'x

In order to collect NMR data to assign the backbone of **b'x**, a $^{15}N/^{13}C$ isotopically labelled sample was required. The **b'x** domain construct was grown as described in section 2.2.7. Two 400mL minimal medium cultures were grown in two 2 Litre flasks. 2g/L of ^{13}C glucose was used, supplemented with ^{15}N-ammonium sulfate and induced for 4 hours. The sample was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1. The purification procedure was the same as that carried out for the ^{15}N labelled **b'x** sample described in the previous section. The protein behaved similarly and gave similar chromatograms and PAGE results. The $^{15}N/^{13}C$ sample was also fractionated into two pools, monomer and dimer.

The concentration by measurement of absorbance at 280nm was calculated to be 27.0 mg, 1.5 mM in a 300 µL sample for the monomer pool. For the dimer pool the concentration was calculated, using the same method, 15.6 mg, 0.8 mM in a 300 µL sample. A small sample of monomer and dimer was buffer exchanged into 10 mM ammonium bicarbonate and analysed by mass spectrometry.

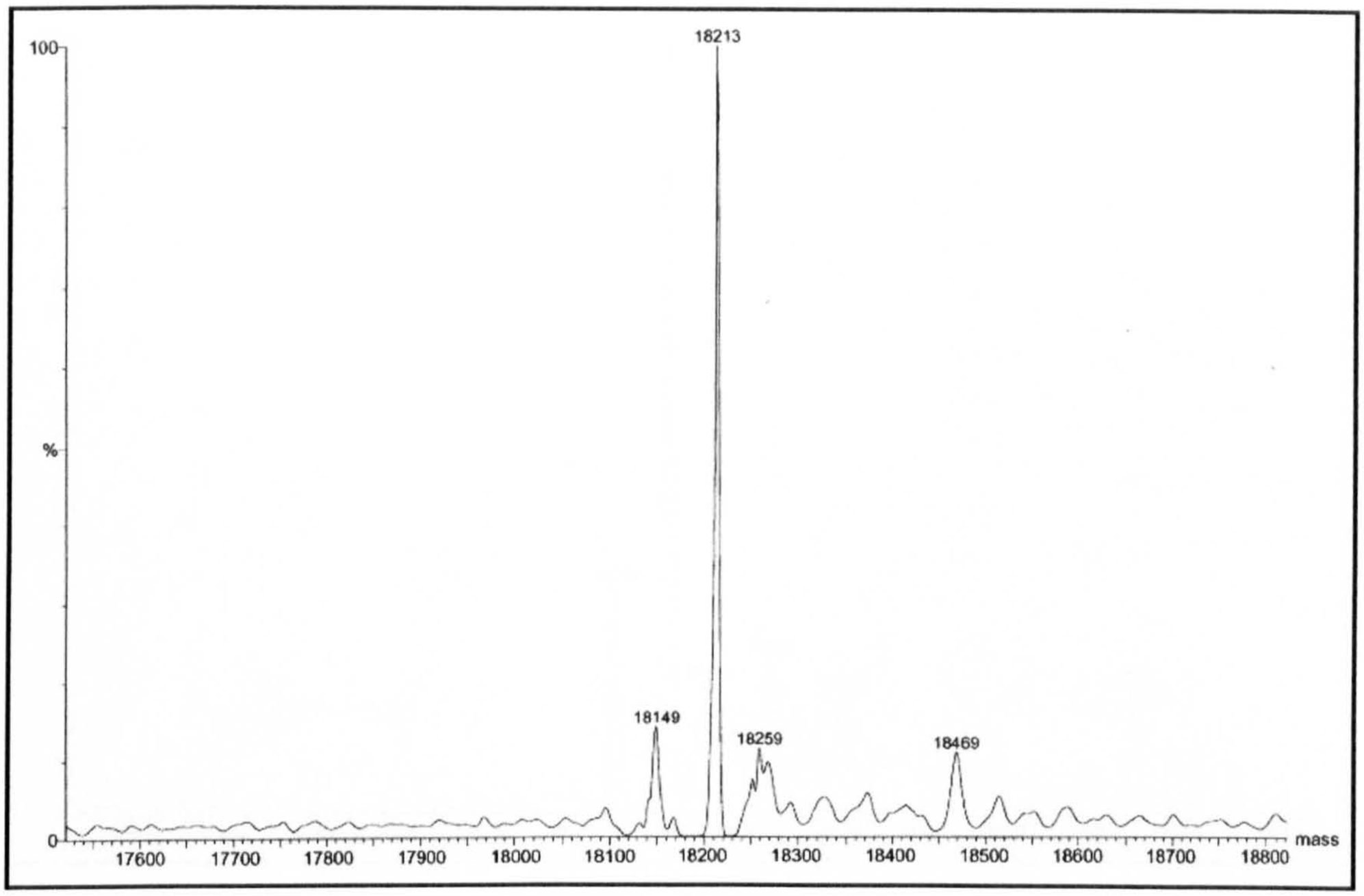

Figure 3-48. Positive ion mass spectrum of $^{13}C/^{15}N$ b'x monomer.

The expected molecular weight if all 205 Nitrogen and all 784 Carbon atoms were labelled with the ^{15}N and ^{13}C isotopes would be 18272 Da since the unlabelled **b'x** domains is 17282 Da. The actual molecular weight recorded electrospray mass spectrometry was 18213 Da, which indicates 931 labelled atoms of the 989 total possible atoms available for labelling and so a 94% labelling efficiency for monomer pool. The 18469 Da species has an extra mass of 256 Da resulting from the His-tag modification. The species with a mass of 18259 Da has an extra mass of 46 Da and so was likely to result from two sodium ions adduct. The minus 64 Da species, at 18149 Da was likely to be product of degradation.

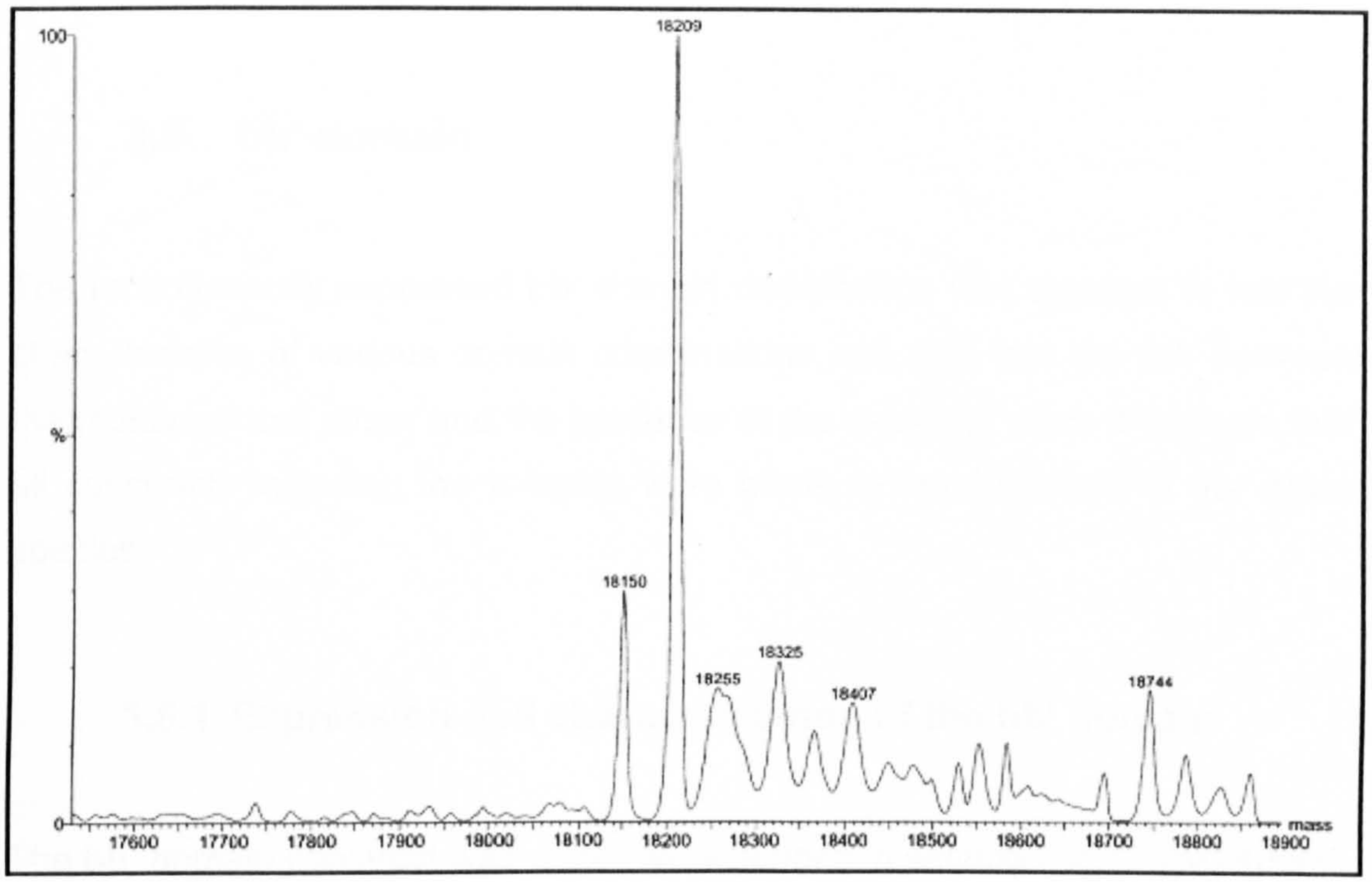

Figure 3-49. Positive ion mass spectrum of $^{13}C/^{15}N$ b'x dimer.

The molecular weight recorded by electrospray mass spectrometry for the dimer species was 18209 Da, which indicates 927 labelled atoms of the 989 total possible atoms available for labelling and so a 94% labelling efficiency for the dimer pool. Several minor species were also recorded. The dimer sample was known to give poor NMR spectra and so generally further analysis any dimer sample were never carried out, but stored to be used in future analysis if required.

3.5. bb' domain

The recombinantly expressed **bb'** domain combination was required to test the characteristics of various domain combinations and also test the link between the monomer and dimer and the presence of the **x**-region; since it appears that all constructs including the **x**-region were prone to the formation of the dimer species.

3.5.1. Expression and characterisation of the bb' domain

The **bb'** domain construct was grown as described in section 2.2.7. Two 400mL LB cultures were grown in two 2 Litre flasks. The sample was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1

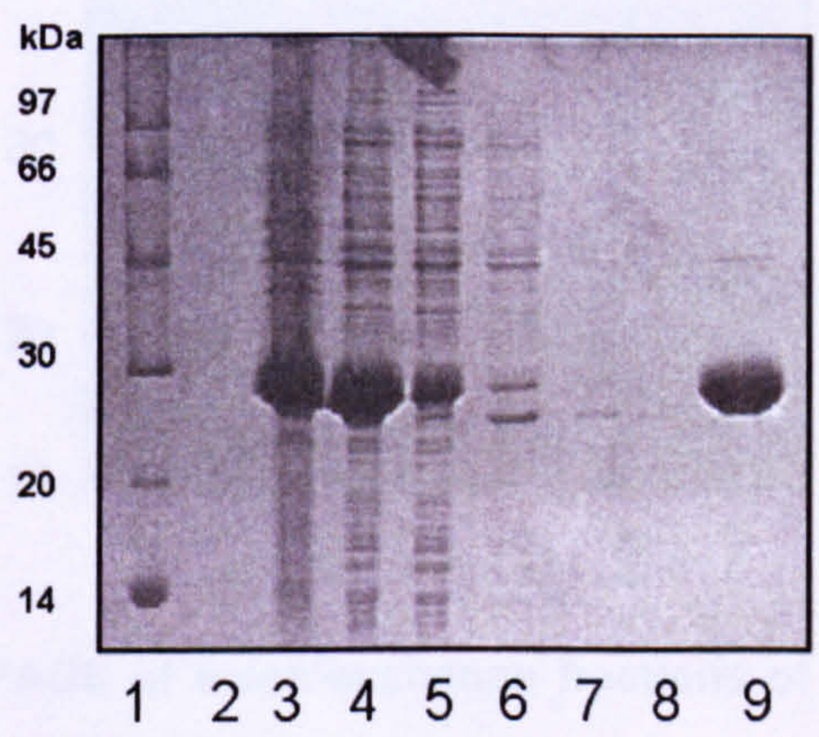

Figure 3-50. bb' domain IMAC purification. Lane 1, low molecular weight marker. Lane 2, empty. Lane 3, total protein. Lane 4, soluble protein. Lane 5, flow through. Lane 6, imidazole wash. Lane 7, low salt wash. Lane 8, empty. Lane 9, hexa-histidine tagged bb' protein elute.

From Figure 3-50, it is clear that purification step was very effective at removing the majority of non-tagged proteins. But the elute still has a protein impurity remaining. So the elute was then prepared and further purified on an anion-exchange column as described in section 2.3.2.

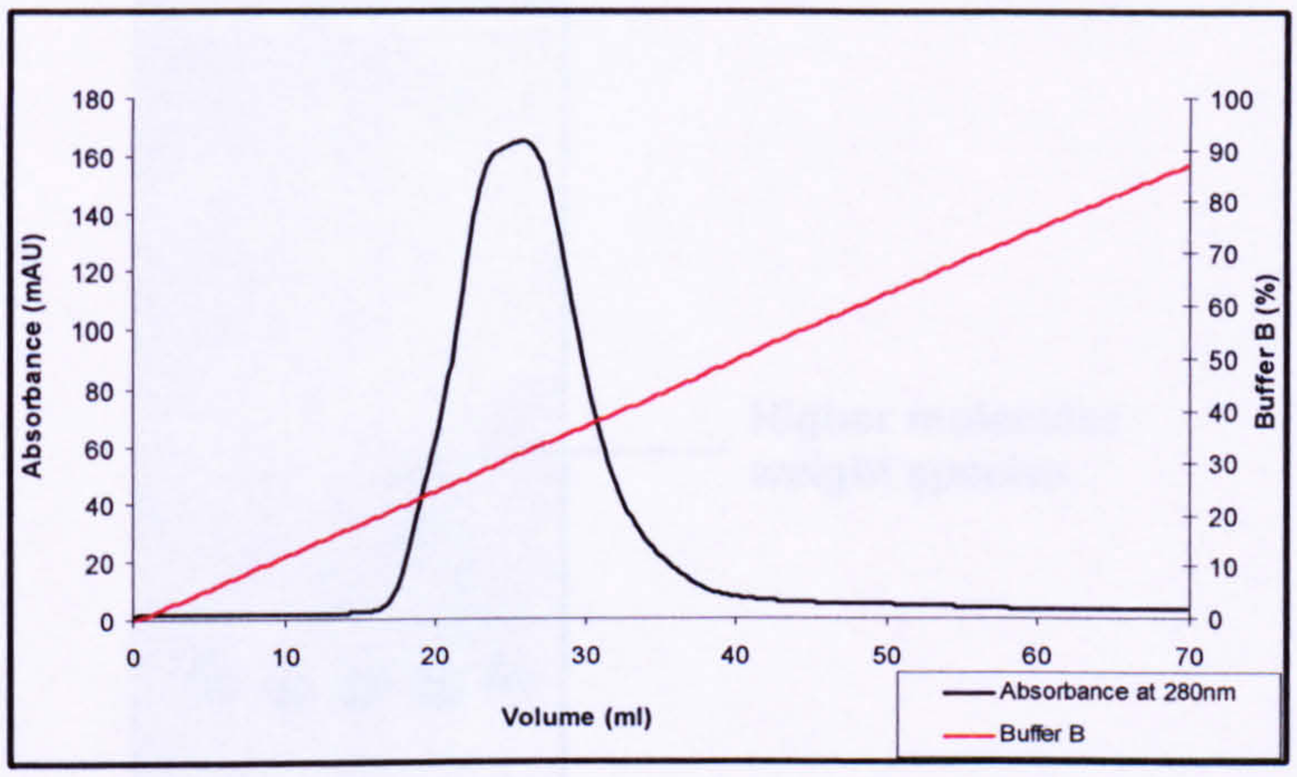

Figure 3-51. Chromatogram of Source 30Q purification of bb'.

To investigate if the peak in the chromatogram was not overlapped with impurities, samples were taken across the peak and run on SDS-PAGE.

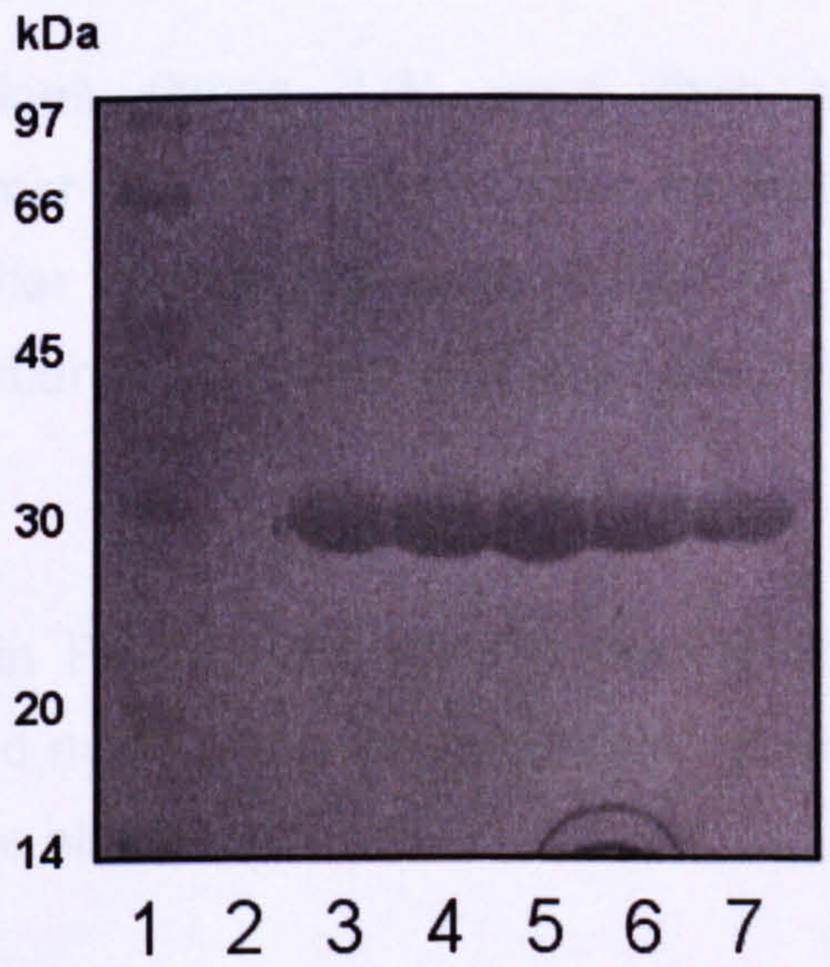

Figure 3-52. 16 % SDS-PAGE of anion-exchange fractions of bb' purification in Figure 3-51.

It was clear from Figure 3-52 that all the fractions taken across the peak shown in Figure 3-51 are homogenous and pure, with no protein contaminates. The same samples were then run on native-PAGE. As shown in Figure 3-53, this analysis revealed a very small amount of the higher molecular weight impurity in lanes 4 and 5, these fractions were therefore discarded.

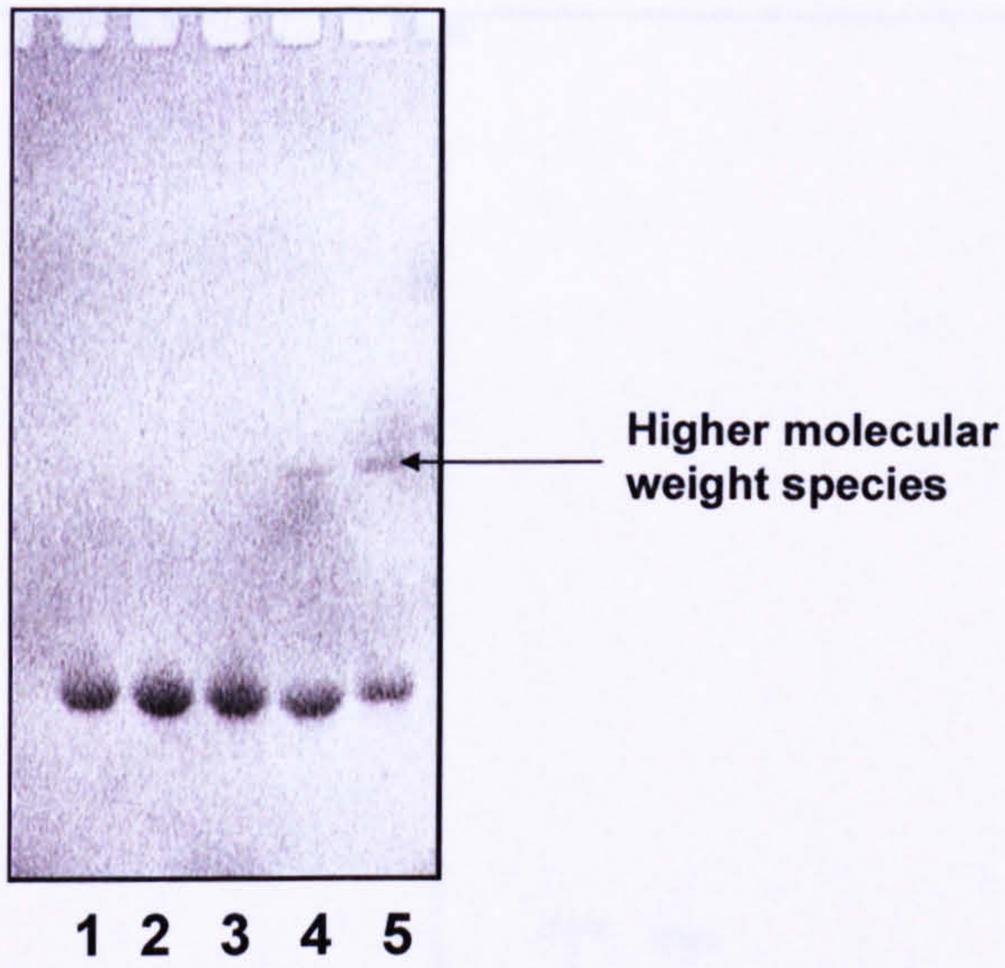

Figure 3-53. 16% native-PAGE analysis of fraction taken across the anion exchange chromatogram in Figure 3-51.

The remaining fractions (lanes 1-3) were then pooled to produce a homogeneous monomer **bb'** sample. The monomer sample was then concentrated and buffer exchanged in to Buffer A. The concentration by measurement of absorbance at 280nm was calculated to be 146 mg, 5.8 mM in a 300 µL sample.

The mass spectrum in Figure 3-54 reveals the 25175 Da species matches exactly to the predicted mass using the sequence. A number of minor species as described earlier are also seen.

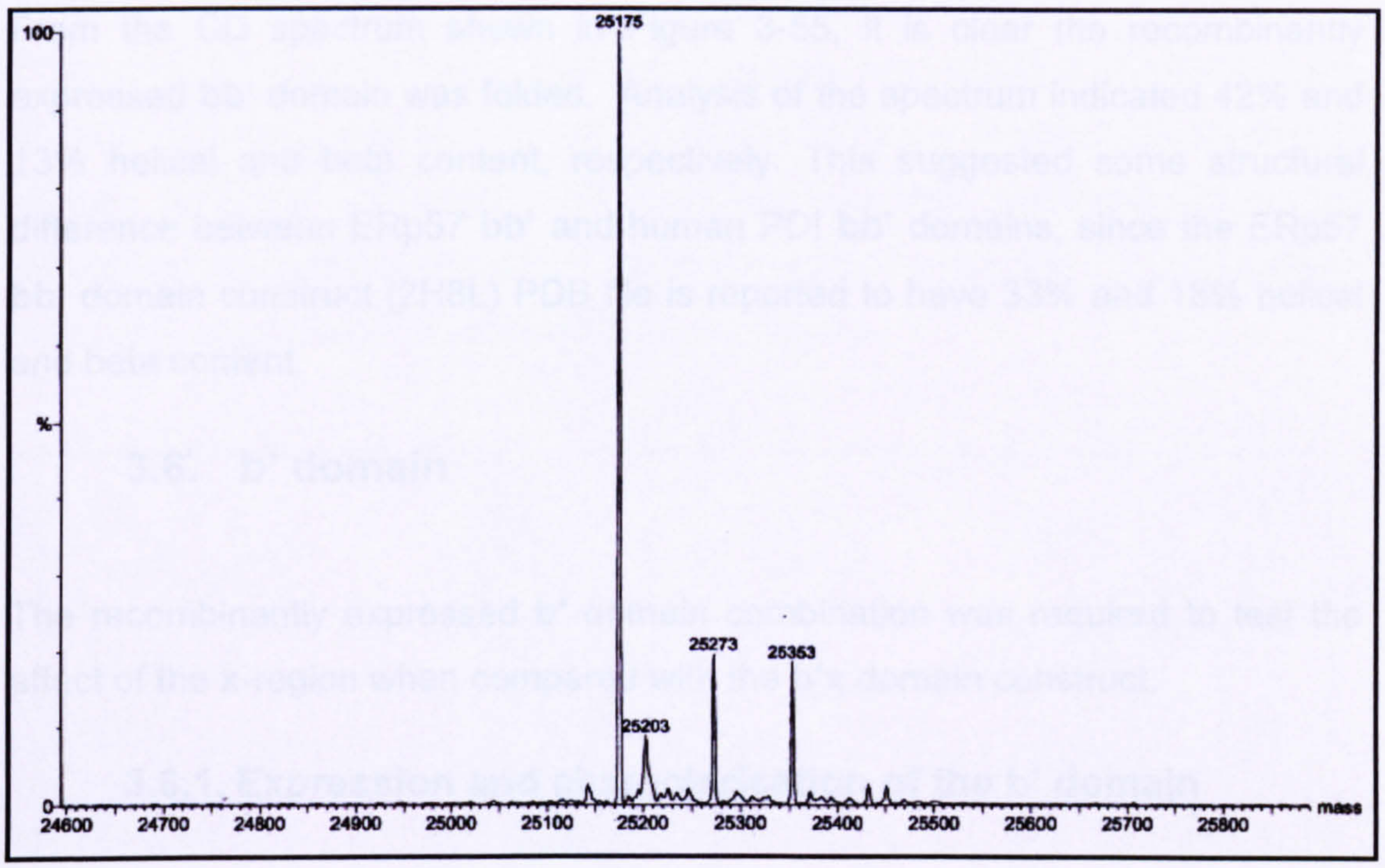

Figure 3-54. Positive ion mass spectrum of bb' monomer sample.

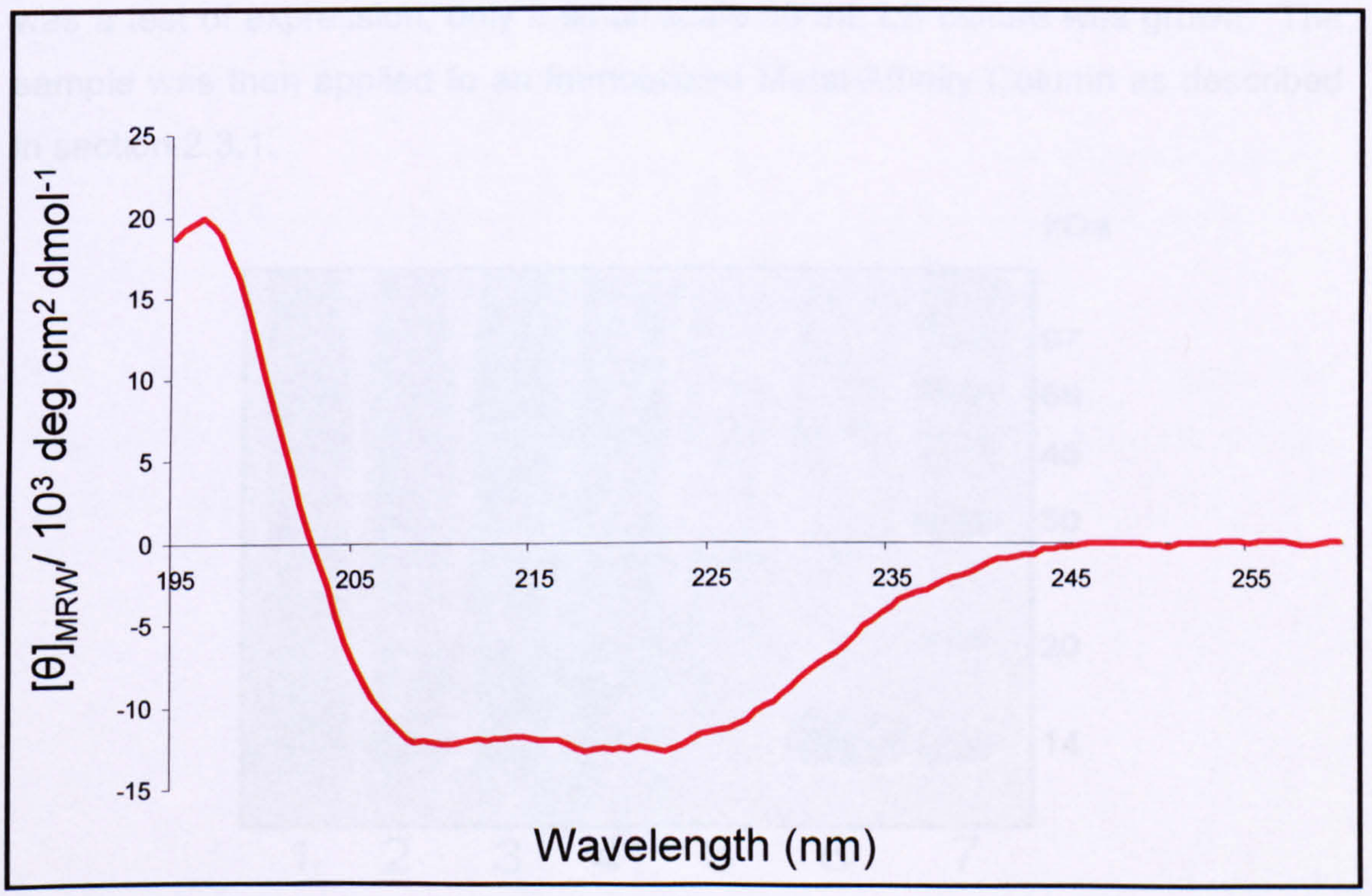

Figure 3-55. A Far UV CD spectrum for the bb' domain. Using a sample, of 2.5uM, 0.06 mg/ml.

From the CD spectrum shown in Figure 3-55, it is clear the recombinantly expressed **bb'** domain was folded. Analysis of the spectrum indicated 42% and 13% helical and beta content, respectively. This suggested some structural difference between ERp57 **bb'** and human PDI **bb'** domains, since the ERp57 **bb'** domain construct (2H8L) PDB file is reported to have 33% and 18% helical and beta content.

3.6. b' domain

The recombinantly expressed **b'** domain combination was required to test the affect of the **x**-region when compared with the **b'x** domain construct.

3.6.1. Expression and characterisation of the b' domain

The **b'** domain construct was grown as described in section 2.2.7. Since this was a test of expression, only a small scale 50 mL LB culture was grown. The sample was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1.

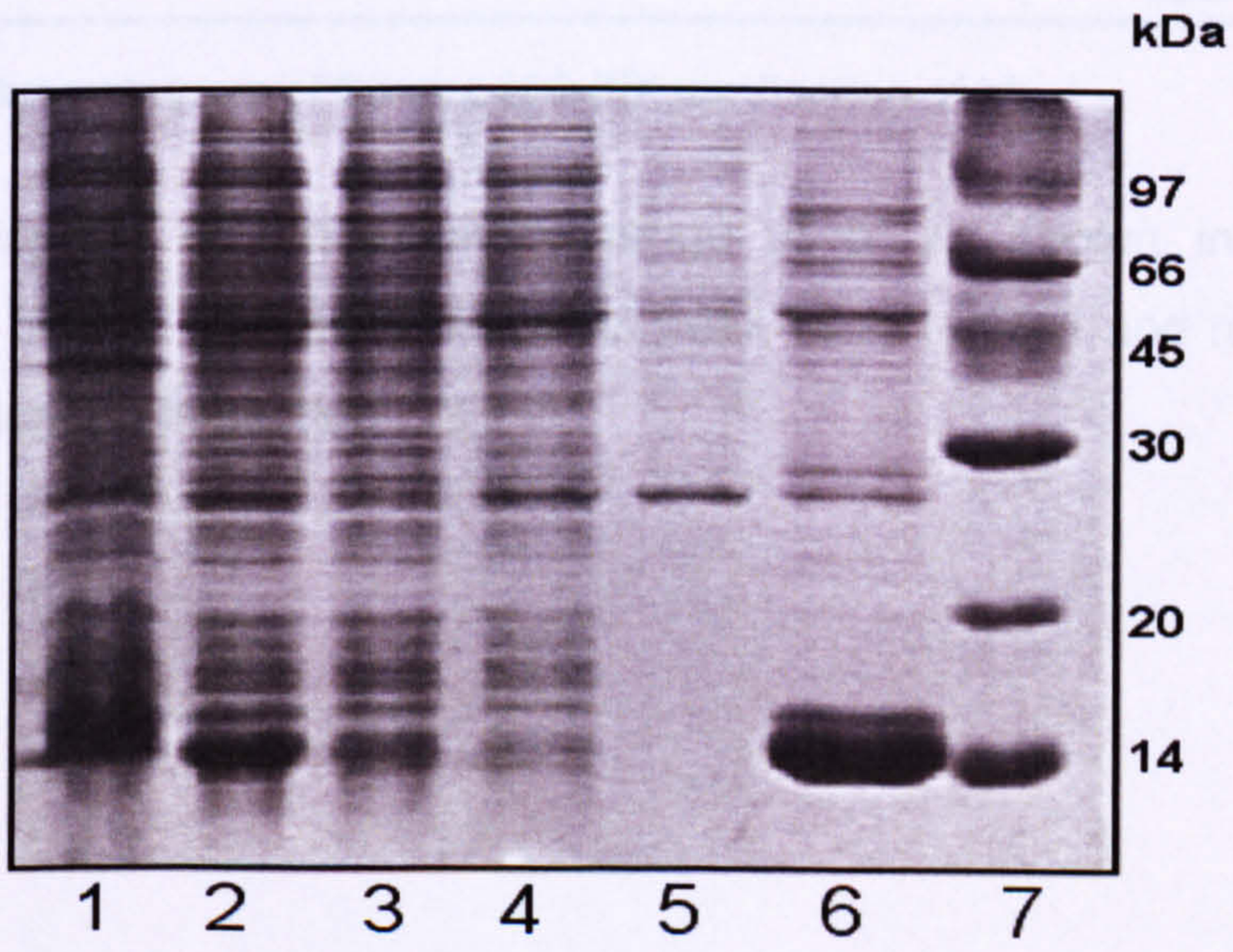

Figure 3-56. b' domain IMAC purification. Lane 1, total protein. Lane 2, soluble protein. Lane 3, flow through. Lane 4, imidazole wash. Lane 5, low salt wash. Lane 6, hexa-histidine tagged b' protein elute. Lane 7, low molecular weight marker.

From IMAC, it is clear that purification step was very effective at removing the majority of protein impurities. But the elute still has a protein high number of protein impurities remaining. So the elute was then prepared and further purified on an anion-exchange column as described in section 2.3.2, but eluted over 50 % of the Buffer B and over a 50 mL volume.

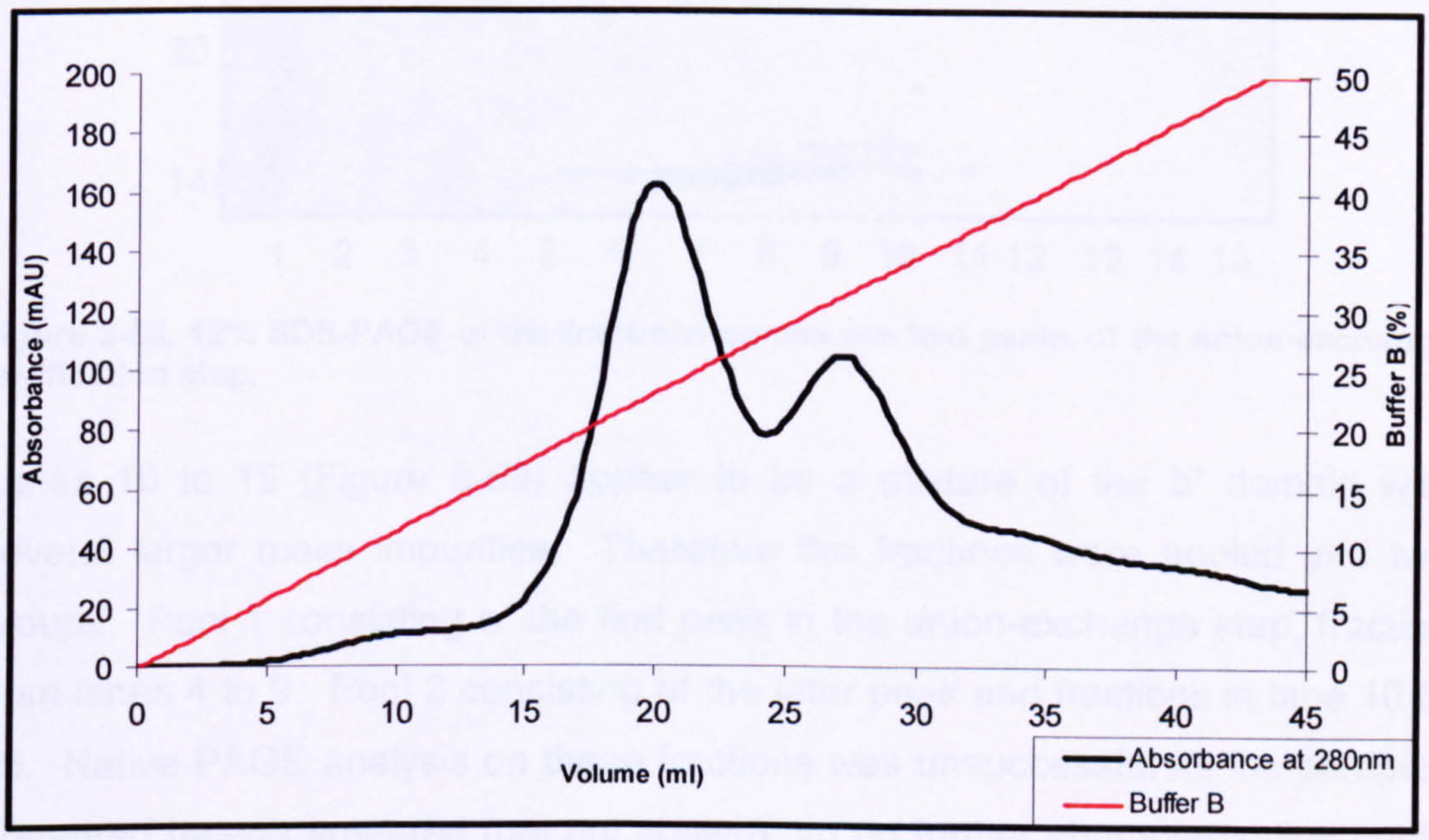

Figure 3-57. Chromatogram of Source 30Q IEX purification of b'.

The chromatogram of the anion-exchange of **b'** as shown in Figure 3-57 revealed two peaks. Fractions from this peak were treated and run on a SDS-PAGE as described in section 2.2.11.

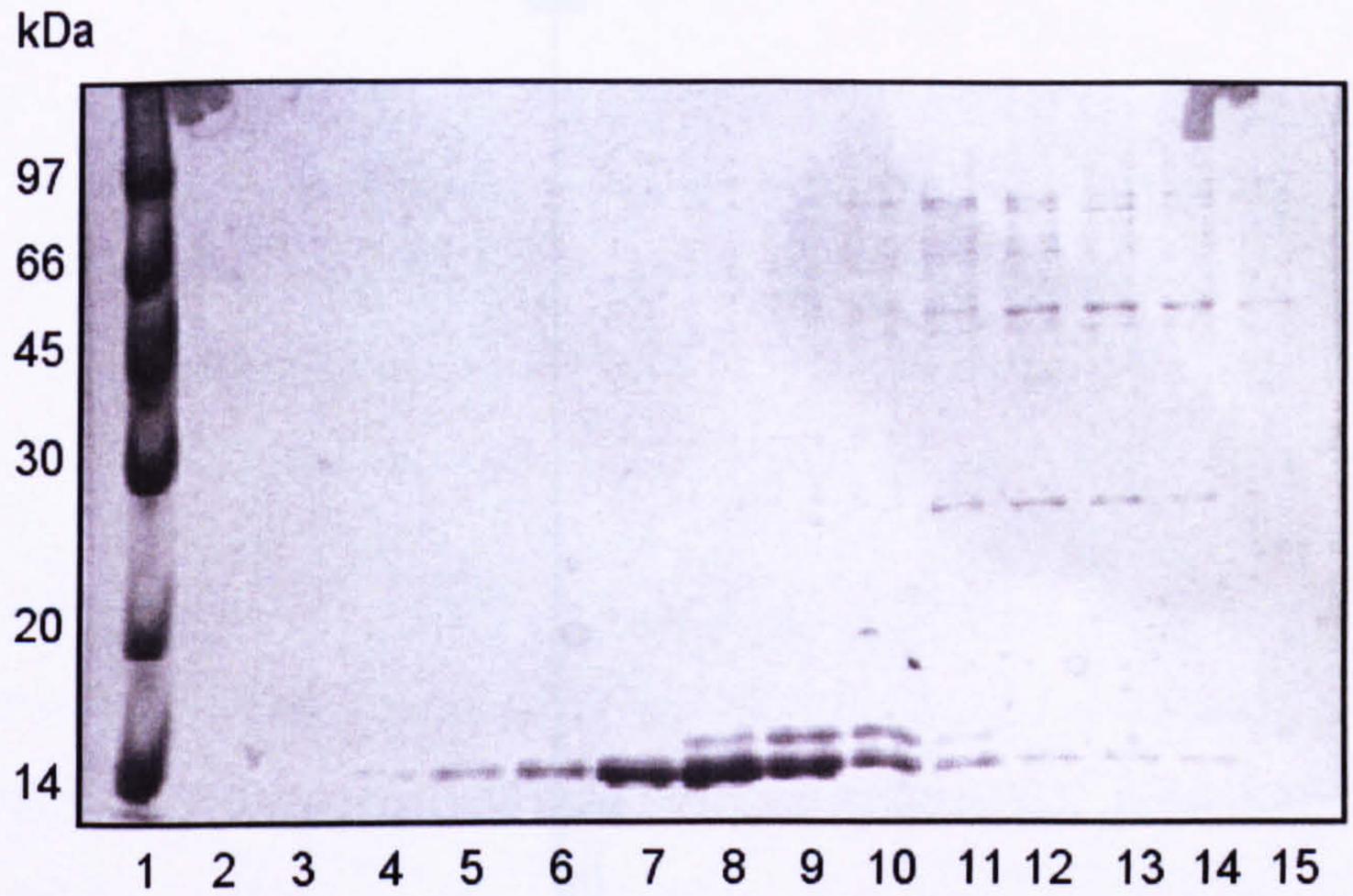

Figure 3-58. 12% SDS-PAGE of the fractions across the two peaks of the anion-exchange purification step.

Lanes 10 to 15 (Figure 3-58) appear to be a mixture of the **b'** domain with several larger mass impurities. Therefore the fractions were pooled into two groups. Pool 1 consisting of the first peak in the anion-exchange step, fraction from lanes 4 to 9. Pool 2 consisting of the later peak and fractions in lane 10 to 15. Native-PAGE analysis on these fractions was unsuccessful as the samples appeared heavily smeared (gel not shown), so no further characterisation could be done.

The concentration by measurement of absorbance at 280nm was calculated to be 4.88 mg, 340 µM in a 500 µL sample for Pool 1. For Pool 2 the concentration was calculated, using the same method, 4.34 mg, 300 µM in a 400 µL sample.

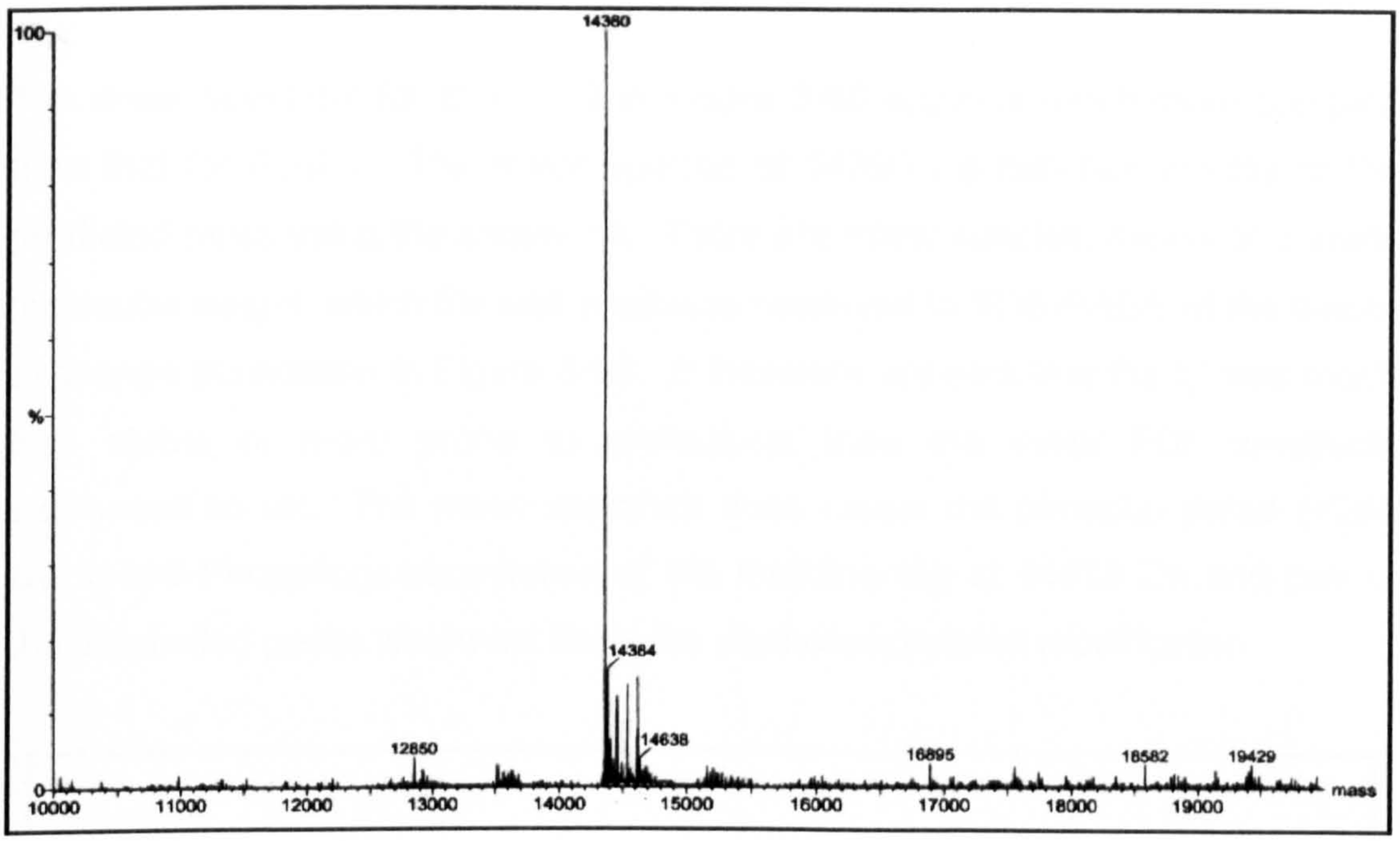

Figure 3-59. Positive ion mass spectrum of b' Pool 1.

The mass spectrum in Figure 3-59 reveals the major 14360 Da species which matches exactly to the predicted mass using the sequence. The spectrum also shows several very minor species discussed in other spectra and a degradation product with a mass of 12850 Da.

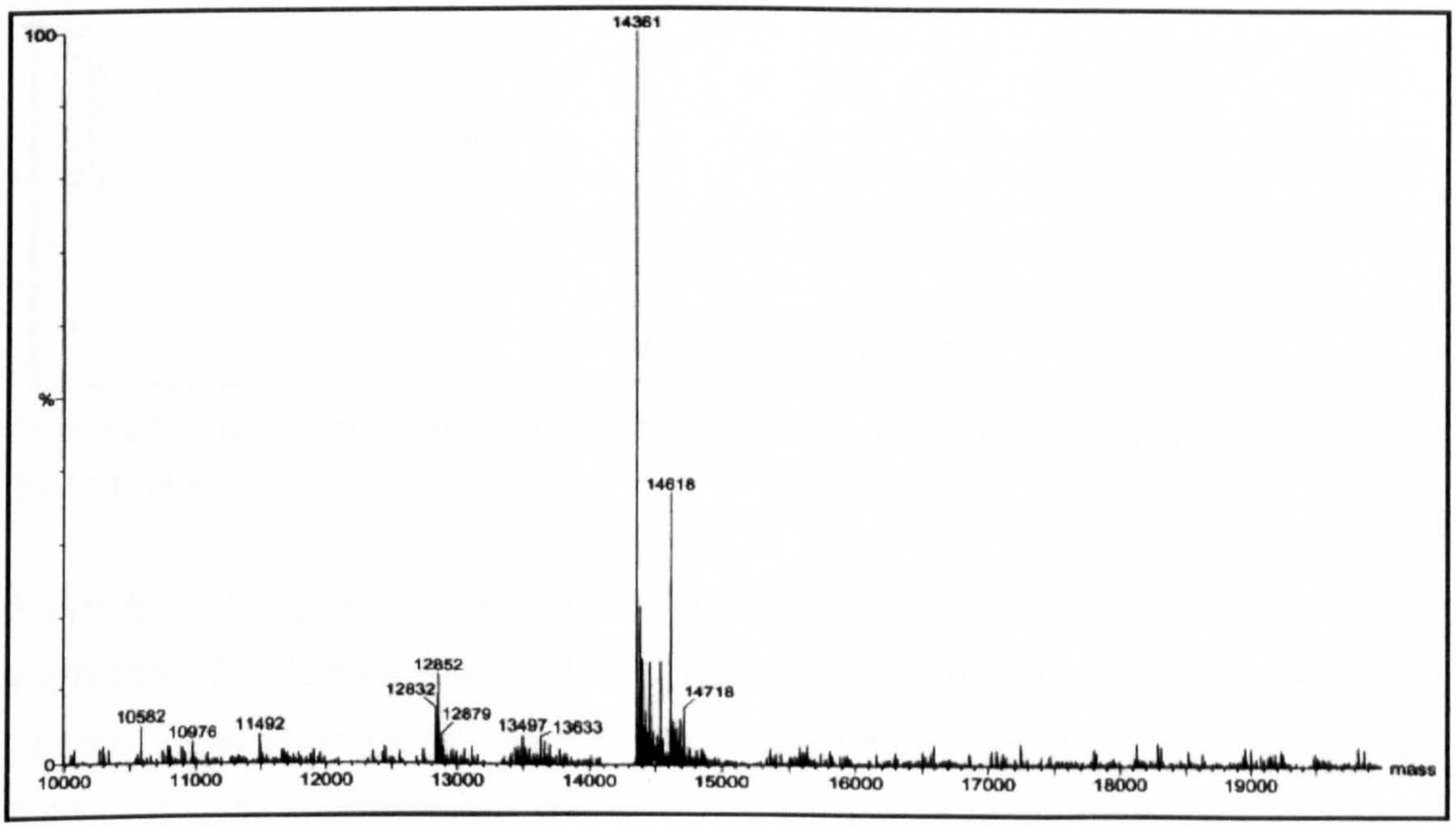

Figure 3-60. Positive ion mass spectrum of b' Pool 2.

The mass spectrum for **b'** Pool 2 in Figure 3-60 appears much more complex than that for Pool 1. The major species of 14360 Da matches exactly to the predicted mass using the sequence. There are many species, mainly of a lower molecular weight, which fits with what was observed in SDS-PAGE of the anion-exchange purification in Figure 3-58. It therefore appears that the **b'** was much less stable or more prone to proteolysis than the other PDI constructs expressed so far. The mass spectrum does reveal the phosphorylated (+258 Da) α-N-6-Phosphogluconoylation of the histidine tag at 14618 Da and one of the unlabelled peaks was most likely the dephosphorylated modification.

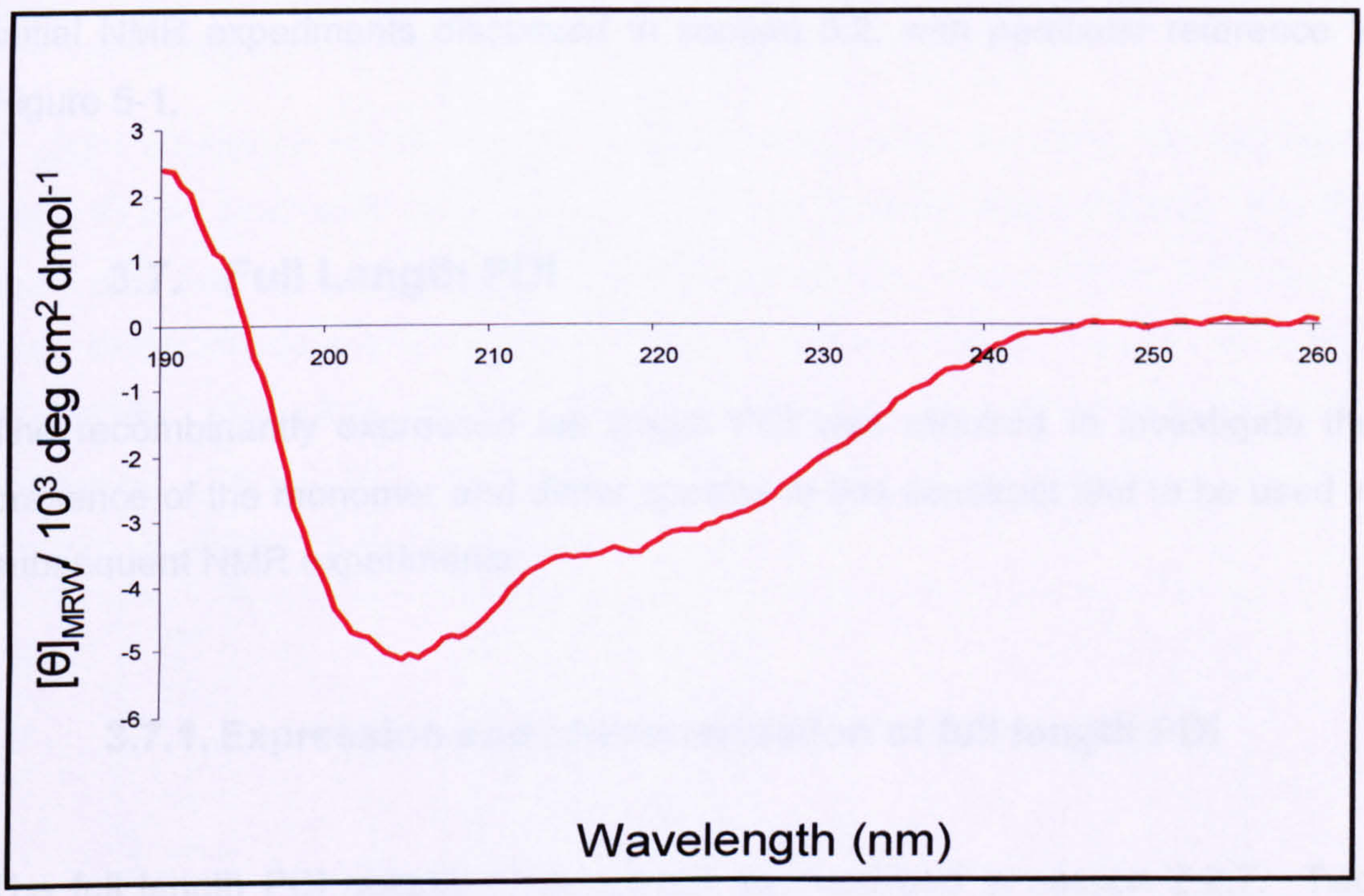

Figure 3-61. A Far UV CD spectrum for the b' domain monomer pool. Using a sample, of 10uM, 0.14 mg/ml.

From the CD spectrum shown in Figure 3-55, it is clear the recombinantly expressed **b'** domain was not well-folded, there appears to be a reduced positive peak at approximately 190nm; larger peak is seen for proteins with a helical content. Furthermore, the two negative peaks at approximately 209 and

222 nm, typical of proteins with beta strand content are shifted toward a lower wavelength. As shown in Figure 2-4, the signal derived from random coils have a minima around 200nm, it appears that this is influencing the spectrum, resulting in the shifted beta strand signal and a reduced helical signal since the random coil signal also has a negative peak around the 190nm range, swamping the helical peak.

Analysis of the spectrum indicated 8% and 41% helical and beta content, respectively. The low helical content recorded suggests a poorly folded domain. The poor folded state of the individual **b'** domain observed here by CD could be indicative of the intrinsic flexibility or conformational heterogeneity also seen in initial NMR experiments discussed in section 5.2, with particular reference to Figure 5-1.

3.7. Full Length PDI

The recombinantly expressed full length PDI was required to investigate the presence of the monomer and dimer species in this construct and to be used in subsequent NMR experiments.

3.7.1. Expression and characterisation of full length PDI

The full length PDI construct was grown as described in section 2.2.7. Two 400mL LB cultures were grown in two 2 Litre flasks. The clarified cell lysate was then applied to an Immobilized Metal-Affinity Column as described in section 2.3.1.

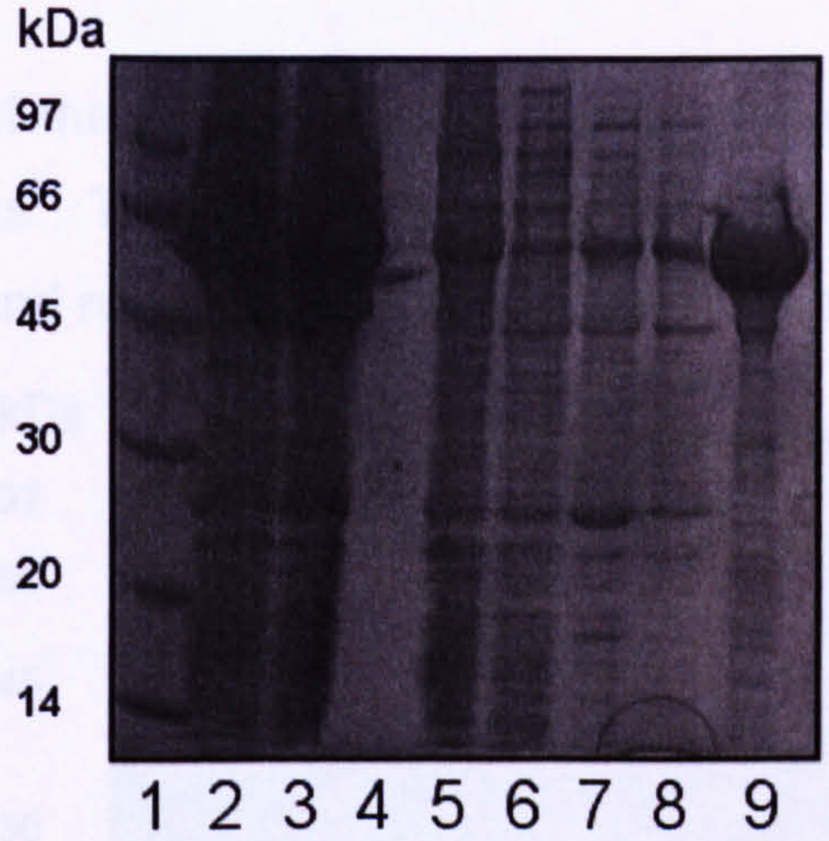

Figure 3-62. 16% SDS-PAGE of the IMAC purification of Full length PDI. Lane 1, low molecular weight marker. Lane 2, total protein. Lane 3, soluble protein. Lane 4, empty lane. Lane 5, flow through. Lane 6, low salt wash. Lane 7, imidazole wash. Lane 8, low salt wash. Lane 9, hexa-histidine tagged Full length PDI protein elute.

A slight difference in the normal protocol here is that, an extra low salt wash (lane 6) was inadvertently added after the flow through was collected. But this extra step appears to have had no ill-desired effects. From Figure 3-62, it was clear that purification step was very effective at removing the majority of protein impurities. The elute sample still has a large number of protein impurities remaining. So the elute was then prepared and further purified on an anion-exchange column as described in section 2.3.2.

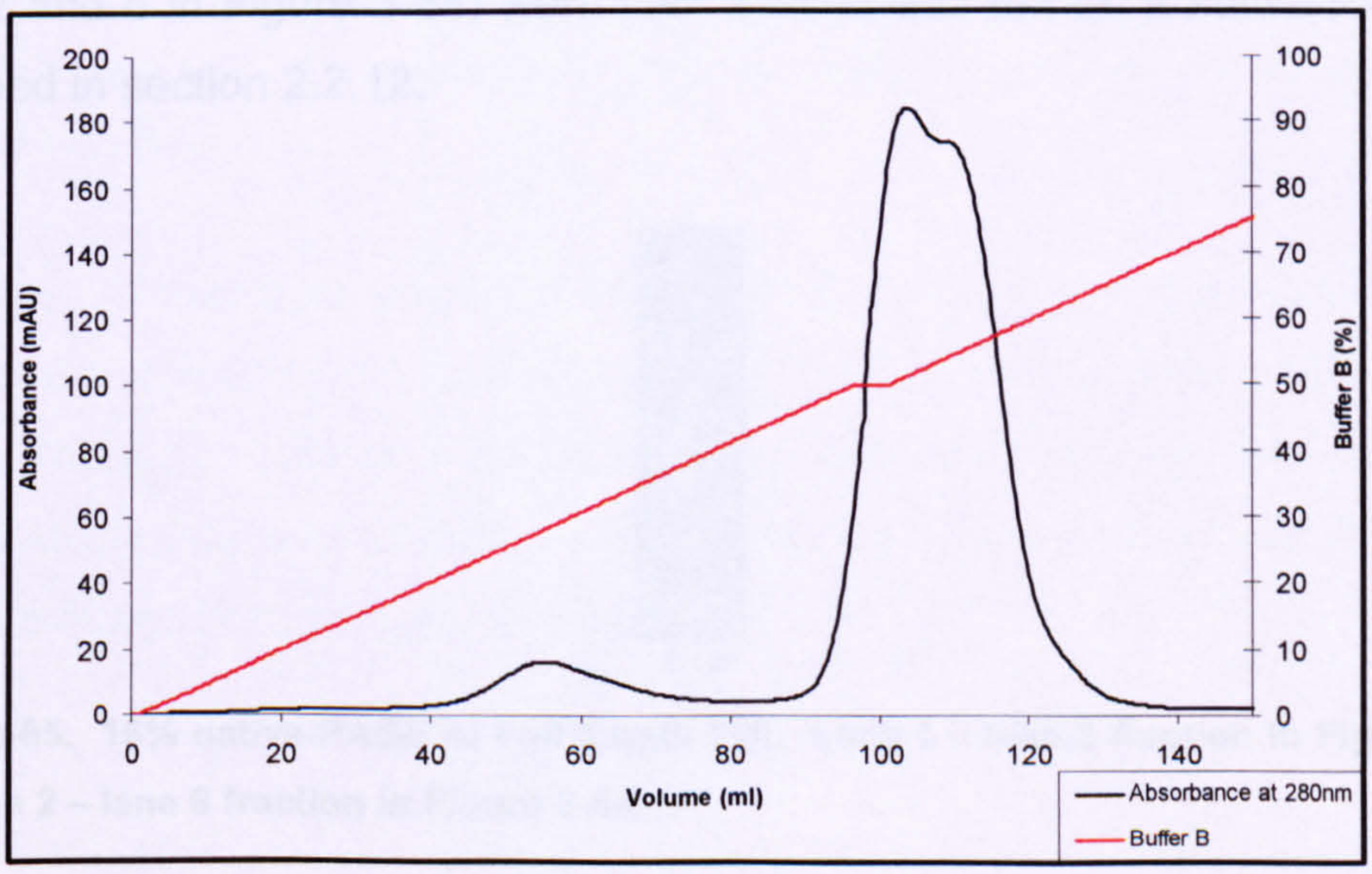

Figure 3-63. Full length PDI Source 30Q purification.

The chromatogram of the anion-exchange of PDI shown in Figure 3-63 reveals two overlapped peaks. To analyse the two peaks further, fractions from these peaks were treated and run on a SDS-PAGE as described in section 2.2.11.

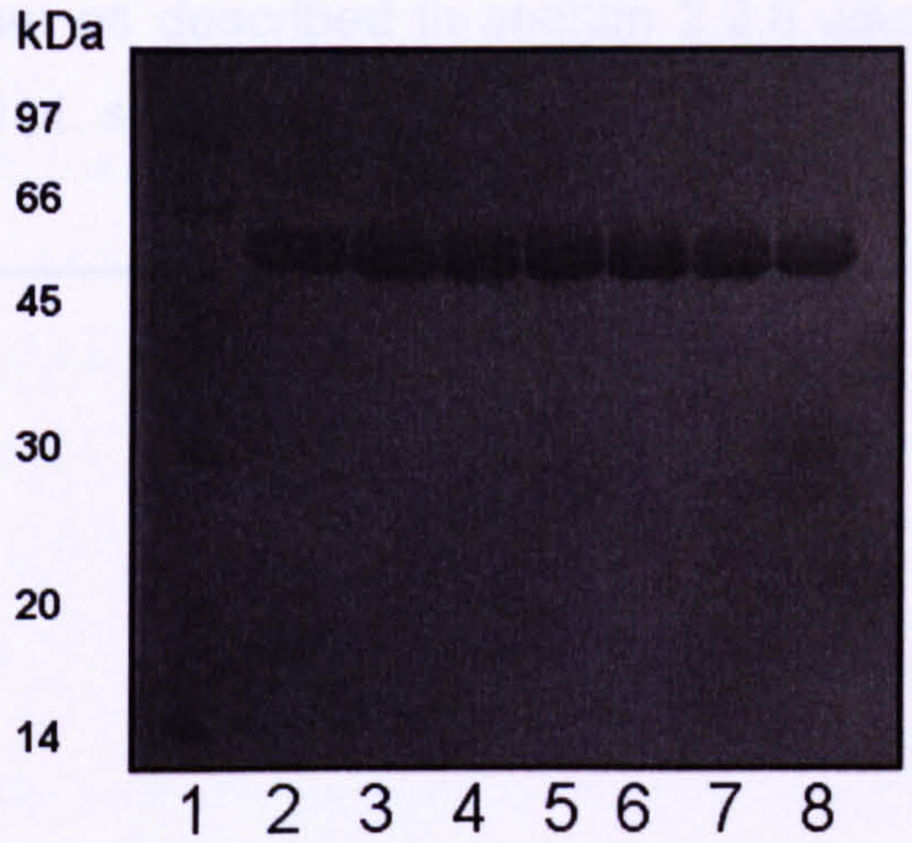

Figure 3-64. 16% SDS-PAGE of fractions taken across peak of full length PDI Source 30Q purification from approximately 100-114mL shown in Figure 3-63.

The fractions across the anion-exchange purification peak run on a 16% SDS-PAGE in Figure 3-64 revealed no impurities. So one fraction sample from each of the overlapping peaks in Figure 3-63 (from fractions at 100mL and 114mL, lanes 2 and 8 in Figure 3-64) were then treated and run on a native-PAGE as described in section 2.2.12.

Figure 3-65. 16% native-PAGE of Full length PDI. Lane 1 – lane 2 fraction in Figure 3-64 and Lane 2 – lane 8 fraction in Figure 3-64.

The two samples taken from each peak in Figure 3-63 appear to have no impurities as shown in Figure 3-64 and under native conditions in Figure 3-65. All the fractions from under the peak in Figure 3-63 were concentrated down and buffer exchanged into Buffer A. The concentration by measurement of absorbance at 280nm as described in section 2.2.8 was calculated to be 312 mg, 5.6 mM in a 500 µL sample.

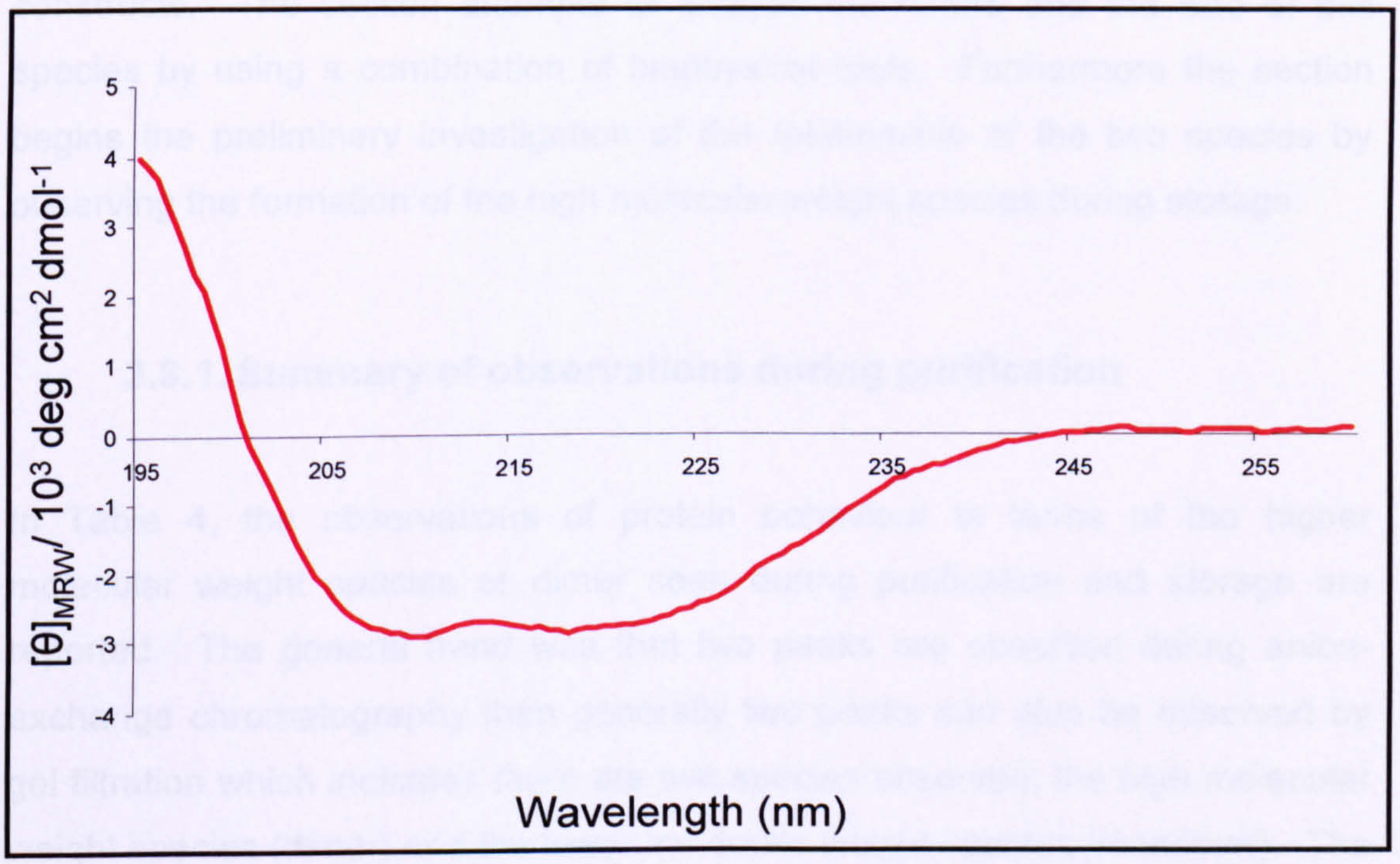

Figure 3-66. A Far UV CD spectrum for the Full length PDI domain. Using a sample of 5uM, 0.28 mg/ml.

From the CD spectrum shown in Figure 3-66, it was clear the recombinantly expressed Full length PDI was well-folded, since there is a large positive peak at approximately 190nm. Furthermore, there are two negative peaks at approximately 209 and 222 nm. Analysis of the spectrum indicated 21% and 37% helical and beta content, respectively.

3.8. Characterisation of the Pool 1 and Pool 2 species identified during purification

This section attempts to answer the question of what is the high molecular weight species which was fractionated into Pool 2 or dimer pool of most protein constructs. The section attempts to analyse the nature and the size of this species by using a combination of biophysical tools. Furthermore the section begins the preliminary investigation of the relationship of the two species by observing the formation of the high molecular weight species during storage.

3.8.1. Summary of observations during purification

In Table 4, the observations of protein behaviour in terms of the higher molecular weight species or dimer seen during purification and storage are reported. The general trend was that two peaks are observed during anion-exchange chromatography then generally two peaks can also be observed by gel filtration which indicates there are two species observed; the high molecular weight species (dimer) and the lower molecular weight species (monomer). The exception was Full length PDI, where two overlapped peaks were observed in the anion-exchange chromatography purification, but only a single species was observed on SDS and native-PAGE.

Note also that the **b'** domain was a very difficult protein sample to work with, prone to aggregation and very unreliable when analysed on PAGE. Although two peaks are seen during both purification steps, the identity of the peaks can only be inferred by what was seen with other domain constructs. The **b'** domain could not be characterised using native-PAGE as the samples smear heavily in the lanes.

So to summarise, with the exception of full length PDI and the **b'** domain all other domain combinations produce two products, a low and high molecular weight species, speculated to be monomer and dimer, from the gel-filtration and native-PAGE analysis but as yet unconfirmed since no molecular weight analysis has been carried out .

Table 4.

PDI domain/ domain combination	Peaks seen in anion-exchange chromatograms	Peaks seen on gel filtration chromatograms	Observed LMW species on native-PAGE (Monomer)	Observed HMW species on native-PAGE (Dimer)	Stored Fractions developed HMW species overtime (Dimer)
b'	2	2	✓	✓*	×
b	1	N/A	✓	×	×
bb'	2	×	✓	✓	✓
Full length PDI	2	×	✓	×	✓
b'x	2	2	✓	✓	✓
bb'x	2	2	✓	✓	✓

Table 4. Table summarising observations during purification and storage of various domain combinations. LMW – low molecular weight, HMW – high molecular weight, N/A – indicating not applicable, ✓ - observed, × - not observed and ✓* - indicating observed in chromatograms but not observed on gels.

3.8.2. Analysis of storage of fractionated samples

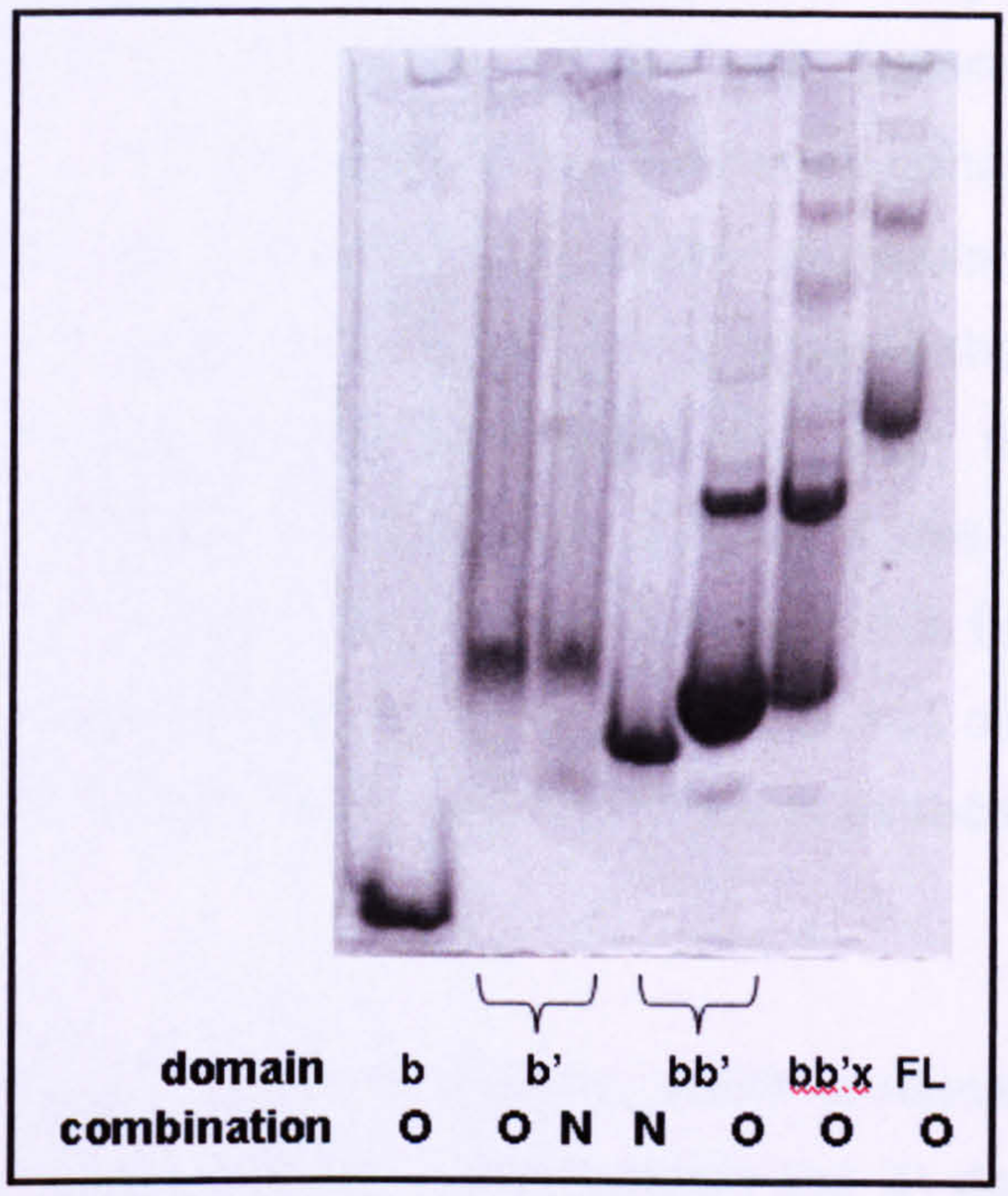

Figure 3-67. 16% native-PAGE of various domain combinations. N – New samples (run on the gel here within two weeks of production), O – Old samples, stored at -12°C in excess of 3 months.

The final column in Table 4 summarises the observation of Figure 3-67, which shows fractionated samples of all domain combinations. Old samples were stored frozen (for a minimum of 3 months storage) and new samples were produced and analysed within two weeks of their production; samples of each were taken and run on native-PAGE. This allowed the analysis of prolonged storage, where samples are stored frozen until required. The trend that appears is that any domain combination which gives two pools during purification, once purified and fractionated, in to the monomer pool will over time produce the high molecular weight species or dimer during storage; this was observed with **bb'**, **bb'x** and **b'x** (**b'x** effect of storage not shown, but was carried out by Dr.K.Wallis).

The interesting exception was full length PDI, which produces two pools during anion-exchange purification, so you would expect to observe the higher molecular weight species or dimer on native-PAGE analysis, this was not the case, but this high molecular weight species was observed after storage as seen in Figure 3-67. The full length PDI protein may behave in similar fashion as **bb'** whereby a very small amount of the higher molecular weight species was produced but the concentration was too low to observe the species on native-PAGE analysis and is observed as it does form over time during storage. Unfortunately since no high molecular weight species was observed on native-PAGE, gel filtration was never done, but in this case this further analysis using gel filtration is expected to identify the existence or not of the high molecular weight species in full length PDI expression which is expected to be formed in a low quantity.

The other unclear case was the **b'** domain, which produces two peaks during anion-exchange and gel filtration. Unfortunately the **b'** domain was prone to aggregation and smears when run on native-PAGE, hence the higher molecular weight species or dimer one would expect to be produced was never observed as highlighted in Table 4, only assumed to exist from the purification chromatograms. Interestingly, after storage at -4°C over a period in excess of three months, the high molecular weight species does not form as reported in Figure 3-67. Hence, the two peaks observed during purification may not actually result from the high molecular weight species or dimer (see section 3.6.1, for details of purification).

3.8.3. Further analysis of the monomer/dimer species by AUC

To further analyse the molecular weight of the monomer and dimer species, analytical ultra centrifugation was used as described in section 2.4.1.

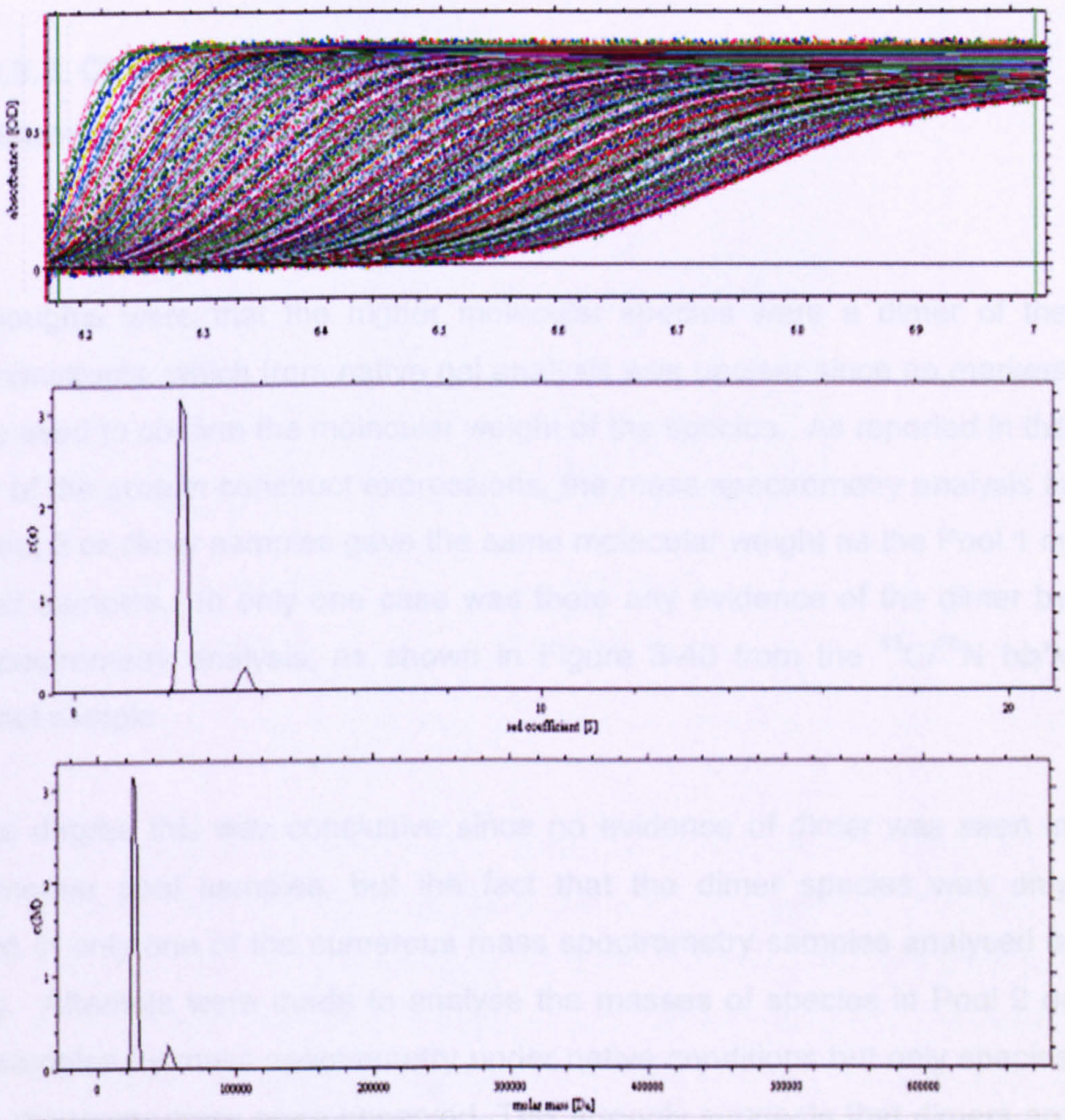

Figure 3-68. Analytical ultra-centrifugation data analysed using SEDFIT and showing sedimentation coefficient and molecular weight for bb'x.

An unlabelled **bb'x** monomer fractionated sample was used for the sedimentation velocity experiments, which produced the chromatograms as shown in Figure 3-68. The monomer species was calculated to be 25 kDa and the dimer 51 kDa. The monomer species is known to be 27.7 kDa, so the calculation by the sedimentation velocity experiments is a little low.

3.8.4. Characterisation of the high molecular weight species or dimer by SDS and native-PAGE

Initial thoughts were that the higher molecular species were a dimer of the protein constructs, which from native gel analysis was unclear since no markers could be used to confirm the molecular weight of the species. As reported in the majority of the protein construct expressions, the mass spectrometry analysis in every Pool 2 or dimer samples gave the same molecular weight as the Pool 1 or monomer samples. In only one case was there any evidence of the dimer by mass spectrometry analysis, as shown in Figure 3-40 from the $^{13}C/^{15}N$ **bb'x** dimer pool sample.

To some degree this was conclusive since no evidence of dimer was seen in any monomer pool samples, but the fact that the dimer species was only observed in only one of the numerous mass spectrometry samples analysed is puzzling. Attempts were made to analyse the masses of species in Pool 2 or 'dimer' samples by mass spectrometry under native conditions but only species with the monomer mass were observed. This strongly suggests that dimers are held together by weak non-covalent forces.

However, since it is known that the **b'** domain contains buried unpaired Cys residues which could, in principle, form intermolecular disulphide bonds, it was

thought essential to check directly for the presence of such bonds. This was necessary since the native-PAGE analyses reported earlier were carried out in non-reducing conditions, whereas SDS-PAGE analyses were conducted in reducing conditions (in the presence of 10% beta-mercaptoethanol).

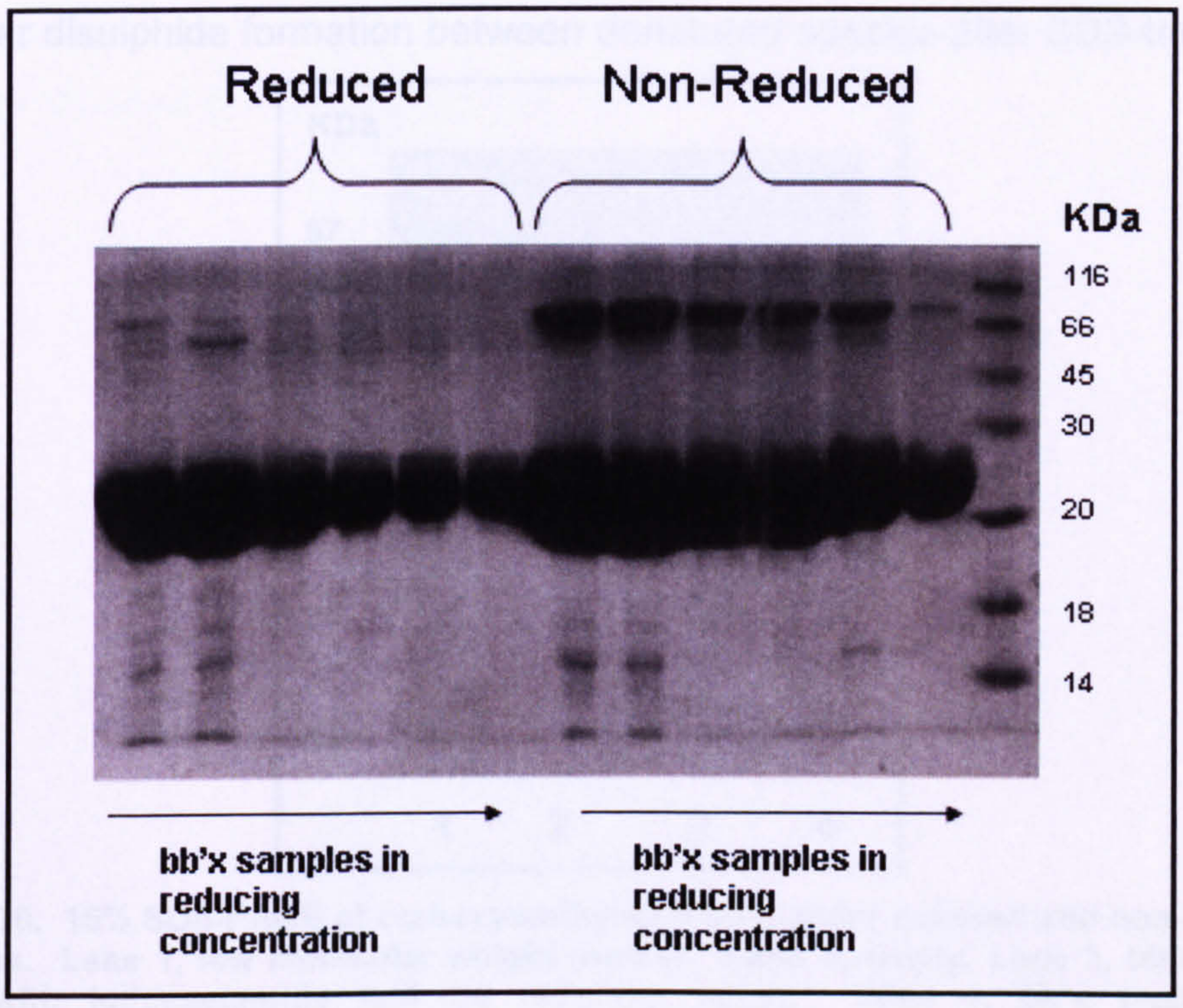

Figure 3-69. 16% SDS-PAGE of bb'x Pool 2 'dimer' of varying concentrations under reduced and no-reduced conditions.

As a first step, samples of **bb'x** Pool 2 'dimer' were prepared for SDS-PAGE either in standard reducing SDS sample buffer, or in presence of SDS but in absence of reducing agent. The results are shown in Figure 3-69 where various protein loadings of reduced and non-reduced samples are compared. Reduced samples show only a single major band at the monomer mass, but non-reduced samples also contain a higher mass band corresponding to dimer, whose intensity increases with sample loading. This band can be assumed to be a disulphide-linked dimer. However, these are heavily overloaded samples; by comparing the monomer and dimer bands in the non-reduced track with the lowest protein loading it is clear that the dimer was a very small fraction of the

bb'x samples (probably <10%). When a comparable sample was analysed by gel-filtration the dimer species constitutes roughly half the material. Hence it seems likely that the dimer observed by these methods was not disulphide-bonded and that the dimer species observed after non-reducing SDS-PAGE in Figure 3-69 was an artefact arising from air oxidation to generate inter-molecular disulphide formation between denatured species after SDS-treatment.

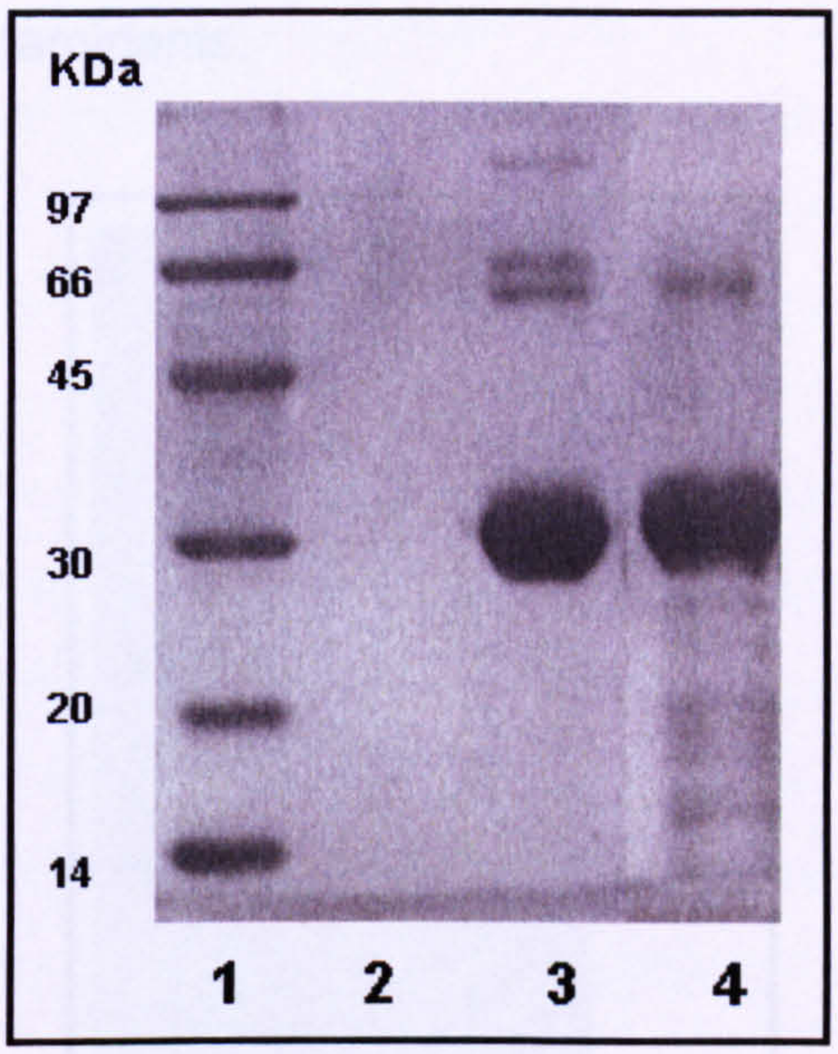

Figure 3-70. 16% SDS-PAGE of carboxymethylated bb'x under reduced and non-reducing conditions. Lane 1, low molecular weight marker. Lane 2, empty. Lane 3, bb'x sample treated with Iodoacetamide and no reducing agent. Lane 4, bb'x treated with iodoacetamide and reducing agent.

To check this interpretation, samples of **bb'x** containing significant levels of dimer were analysed by SDS-PAGE in presence of the thiol-blocking reagent iodoacetamide to block artefactual disulphide bond formation or isomerisation. Figure 3-70 shows the results of SDS-PAGE analysis of a Pool 2, dimer sample of **bb'x** after sample treatment with iodoacetamide in presence of SDS and in absence (lane 3) or presence (lane 4) of beta-mercaptoethanol. The former treatment should retain any pre-existing protein disulphides, whereas the latter should break them and alkylate the free thiol groups generated. As is clear from Figure 3-70, the dominant species observed in both samples is the **bb'x** monomer, confirming that there was no disulphide-linked dimer. Interestingly,

lane 3 shows the presence of three higher molecular weight species; one minor band is common to both lanes 3 and 4 and is likely to be a non-disulphide-bonded contaminant whereas the other two of these are clearly disulphide-linked since they are not observed in lane 4. It is possible that these bands are oligomers of **bb'x**, but since lane 4 also contains a number of lower molecular weight species not seen in lane 3, it is likely that these bands represent disulphide-bonded contaminants.

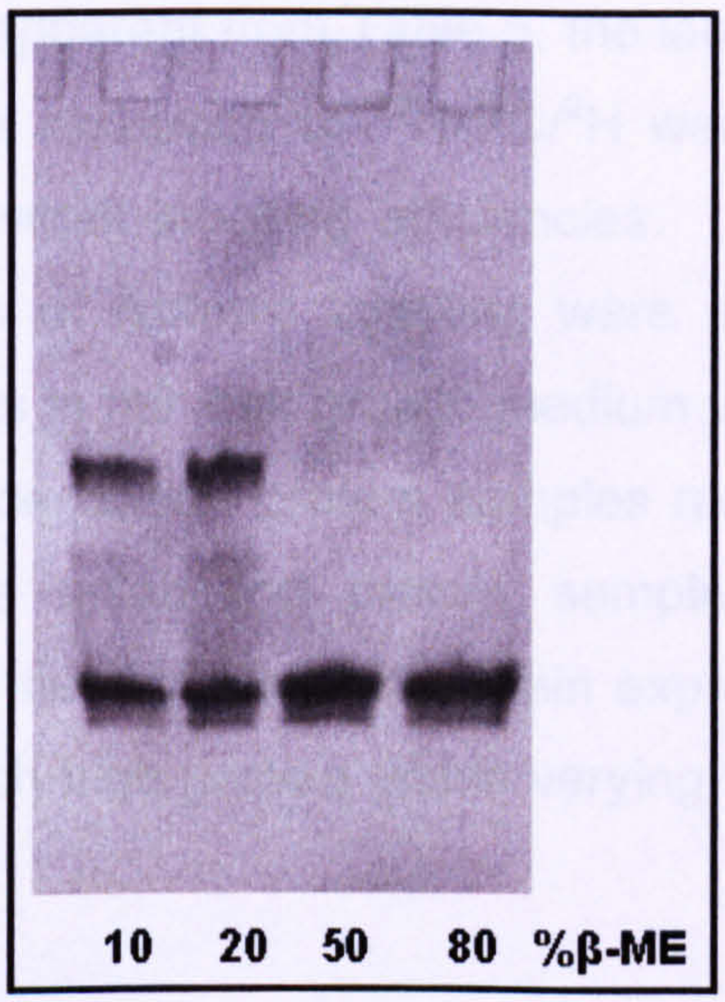

Figure 3-71. 16% native-PAGE of bb'x with different percentage of β-Mercaptoethanol in the loading buffer

As a final check, samples of **bb'x** containing significant amounts of dimer were analysed by native (non-denaturing) PAGE but in the presence of increasing quantities of the reductant beta-mercaptoethanol. As shown in Figure 3-71 this sample showed the presence of roughly equal quantities of monomer and dimer species in presence of 10% or 20% mercaptoethanol. These are strongly reducing conditions implying that the dimer was not formed as a result of covalent disulphide bonds. At extremely high mercaptoethanol concentrations (≥50% by volume) the dimer band disappears, but this is likely to be a solvent denaturation effect rather than the effect of reduction since it is not observed in the lower concentrations which are nevertheless strongly reducing.

3.9. Discussion

As summarised in Table 5 the minimum protein concentration required for NMR analysis were exceeded in almost all cases with the exception of a $^{15}N/^{13}C$ **bb'x** dimer sample, but as described later, monomer samples were used to collect NMR data to enable backbone assignments and dimer samples were not used for NMR analysis. Also apparent from Table 5, the labelling efficiency generally exceeded 96%, with the exception of $^{15}N/^{13}C/^{2}H$ were the lower ^{2}H labelling efficiency reduced the overall labelling efficiencies. This summary highlights the fact that high levels of isotopic labelling were achieved and the use of isotopic labelling reagents in minimal growth medium was adequately optimised to produce the high concentration protein samples required for NMR analysis. Also highlighted by the unlabelled protein samples grown in LB, is the exceptional efficiency of the recombinant protein expression technique used in the research to allow such high protein yields varying from a respectable 28mg to a massive 312mg.

This chapter details and characterises the expression and purification of the various domains and domain combinations to be studied further. It was shown that sufficient quantities for analysis can be produced and more importantly a multi-step purification protocol was established to ensure a homogenous sample can be manufactured, providing the basis for more detailed structural analysis.

It was clear that the initially named high molecular weight species was indeed a dimer. This was conclusively shown by the analytical ultra-centrifugation experiment, even though the identified masses were not exactly as those recorded by mass spectrometry or predicted from sequence analysis. The data from mass spectrometry and gel electrophoresis are all compatible with the interpretation that the 'dimer' form of **bb'x** was a weak non-covalent complex.

The chaperone function analysis was regrettably abandoned as a result of the lack of reproducibility. The initial results were interesting, suggesting the importance of the **b'** domain for chaperone activity and the lack of activity by the **bb'x** domain; but no firm conclusions can be drawn due to the ambiguity of the data. With a more robust method for analysing chaperone activity, many interesting questions could be addressed.

Table 5:

Protein	**Isotopic labelling**	**Amount of protein (mg)**	**NMR sample (mg)**	**Labelling efficiency (%)**
b	Unlabelled	28	N/A	N/A
b	^{15}N	28	18	98.7
bb'x monomer	Unlabelled	99	N/A	N/A
bb'x dimer	Unlabelled	10	N/A	N/A
bb'x monomer	^{15}N	64	41	97
bb'x dimer	^{15}N	57	41	97
bb'x monomer	$^{15}N/^{13}C$	179	41	96
bb'x dimer	$^{15}N/^{13}C$	29	41	96
bb'x monomer	$^{15}N/^{13}C/^{2}H$	60	41	77
bb'x dimer	$^{15}N/^{13}C/^{2}H$	85	41	78
b'x monomer	^{15}N	35	26	97
b'x dimer	^{15}N	18	26	97
b'x monomer	$^{15}N/^{13}C$	27	27	94
b'x dimer	$^{15}N/^{13}C$	16	27	94
bb'	Unlabelled	146	N/A	N/A
b' Pool 1	Unlabelled	80	N/A	N/A
b' Pool 2	Unlabelled	64	N/A	N/A
FL PDI	Unlabelled	312	N/A	N/A

Table 5. Summary table of protein yields from 800mL expression of various domains and domain combinations. Unlabelled proteins were expressed in LB, ^{15}N and $^{15}N/^{13}C$, expressed in minimal medium with appropriate labelled reagents. $^{15}N/^{13}C/^{2}H$ expressed in minimal medium with appropriate labelled reagents and H_2O substituted with 96% D_2O. b' yields are estimated from a 50 mL expression, since no large scale expression was carried out. NMR sample, refers to amount of protein required for a 1.5mM sample, minimum required for NMR experiments.

Chapter 4. Preparation of $^{2}H/^{13}C/^{15}N$-labelled bb'x

4.1. Introduction

A complete backbone assignment from approaches using TOCSY experiments on $^{15}N/^{13}C$ labelled **bb'x** was not possible due to the low resolution of data obtained from NMR triple resonance experiments. It was hoped the two complementary TROSY (Transverse relaxation-optimised spectroscopy) experiments will allow completion of the gaps in the **bb'x** assignments. TROSY experiments require a triple-labelled $^{2}H/^{13}C/^{15}N$ sample. The advantage of TROSY experiments is that deuteration of a protein causes an increase in sensitivity of TROSY experiments in comparison to TOCSY experiments (described in detail in section 1.4.4.3). It was therefore hoped that with the aid of these two experiments the complete backbone could be assigned.

The difficulty lies in cell growth and expression of protein in high levels of deuterium. While most isotopic substitutions (e.g. ^{13}C for ^{12}C and ^{15}N for ^{14}N) have minimal effects on the chemistry of the labelled compounds, it is well known that $^{2}H_2O$ (otherwise known as D_2O) is different in many chemical and physical properties from $^{1}H_2O$ (H_2O). In earlier Chapters, cell growth was not hindered by the presence of ^{15}N and/or ^{13}C; furthermore the incorporation of each isotope appeared to be high. This is not the case for ^{2}H labelling, cell growth is greatly affected by the presence of deuterium. So it is essential that cells are adapted to high levels of deuterium, which was unnecessary for $^{13}C/^{15}N$ labelling. This Chapter describes methods used to adapt cells to high concentrations of deuterium, maximising expression and level of incorporation of triple isotopically labelled protein. Ideally isotopic labelling of between 50-70% is required; at this range the decrease in sensitivity from the limited number

of protons is offset by the reduction in peak linewidths; as discussed previously in section 1.4.6.

4.2. Adaptation of Cells and Expression of Protein in Increasing D_2O Conditions

The first stage to produce any $^{2}H/^{13}C/^{15}N$ would be to adapt cells to increasing amounts of ^{2}H in the growth medium. As described in section 2.2.10 deuterium was supplied by substituting D_2O for H_2O in the minimal medium. Firstly the **bb'x** transformed *E.coli* cells were streaked on LB plates with appropriate antibiotics from which a single colony was used to grow a 20 mL overnight culture. The cells were then centrifuged at 15900 x g for 10 min at 4 °C. The cell pellet was then resuspended in 5 mL minimal medium containing 35% D_2O. The O.D. of the resuspended cells was measured and the appropriate amount added to inoculate a larger 40mL 35% D_2O-containing minimal medium at O.D. 0.1 and this 40 mL culture was grown overnight. A sample of 600 µL of overnight culture was taken, added to 400 µL of 50% glycerol and stored at a temperature of -80°C as a glycerol back-up culture; to be used in future. This process was repeated as shown in Figure 4-1 until cell growth was seen in the maximum amount of D_2O present.

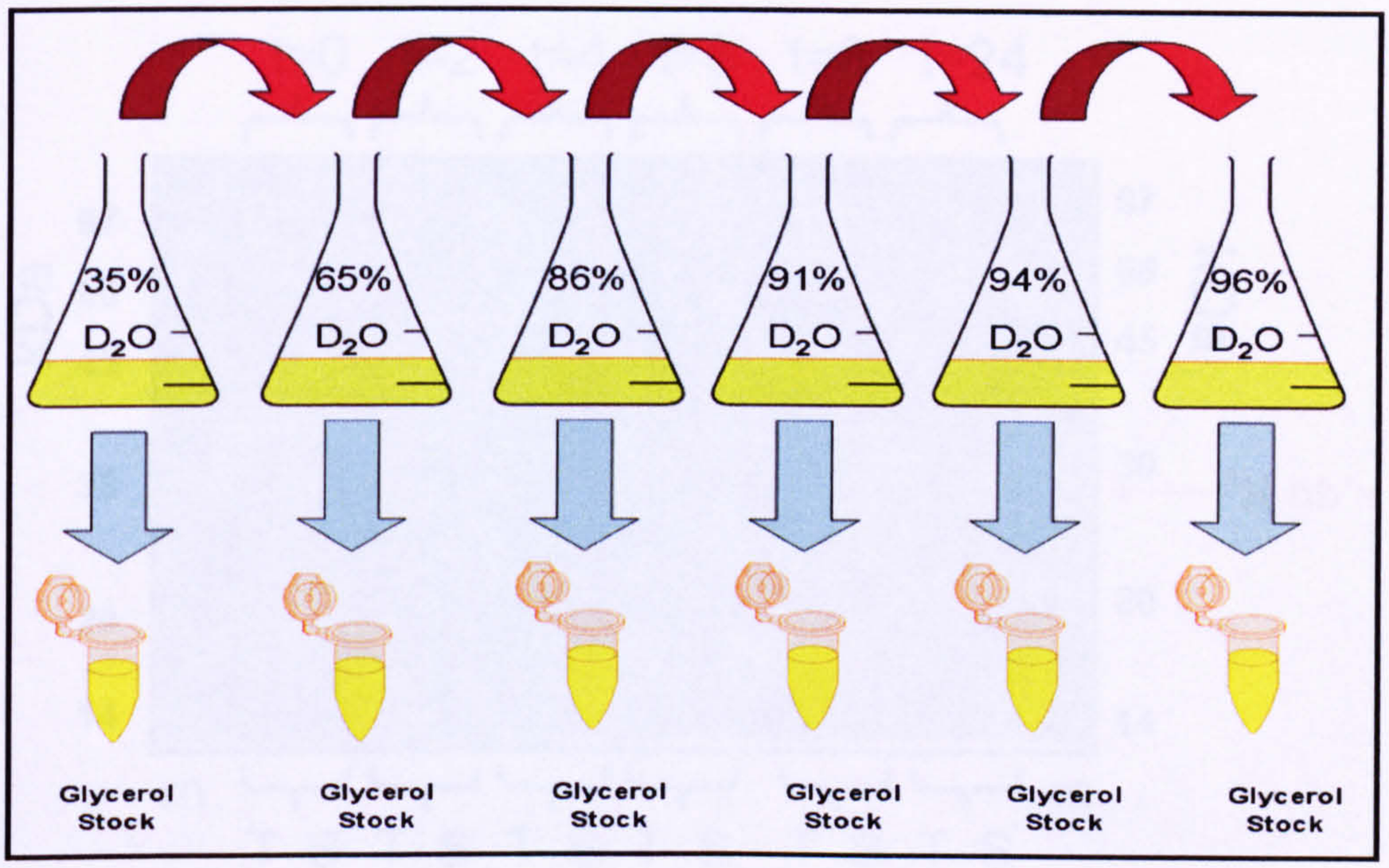

Figure 4-1. Schematic diagram of adaptation of bb'x *E.coli* cells to 96% D_2O.

The adapted cells were able to grow with large step increases of D_2O to 86% after which a jump to 96% resulted in no growth. The growth was measured by increases in cell optical density at a wavelength of 600nm. Smaller increases in D_2O were therefore required in the 90%+ range of D_2O. By this method it was possible to produce cells adapted to growth in 96% D_2O.

4.2.1. Test expression of adapted cells

To test if the cells adapted to 96% D_2O had retained the ability to produce soluble recombinant protein, cells were re-grown using the glycerol stock samples, by first streaking on a LB agar plate with appropriate antibiotics; a colony was transferred to 96% D_2O in minimal medium conditions and samples taken over a 24 hour period of induction. 1 mL samples were taken before induction 0 hours, 2, 4, 6, 8 and 24 hours after induction. The samples were treated as described in Figure 3-6 and run SDS-PAGE.

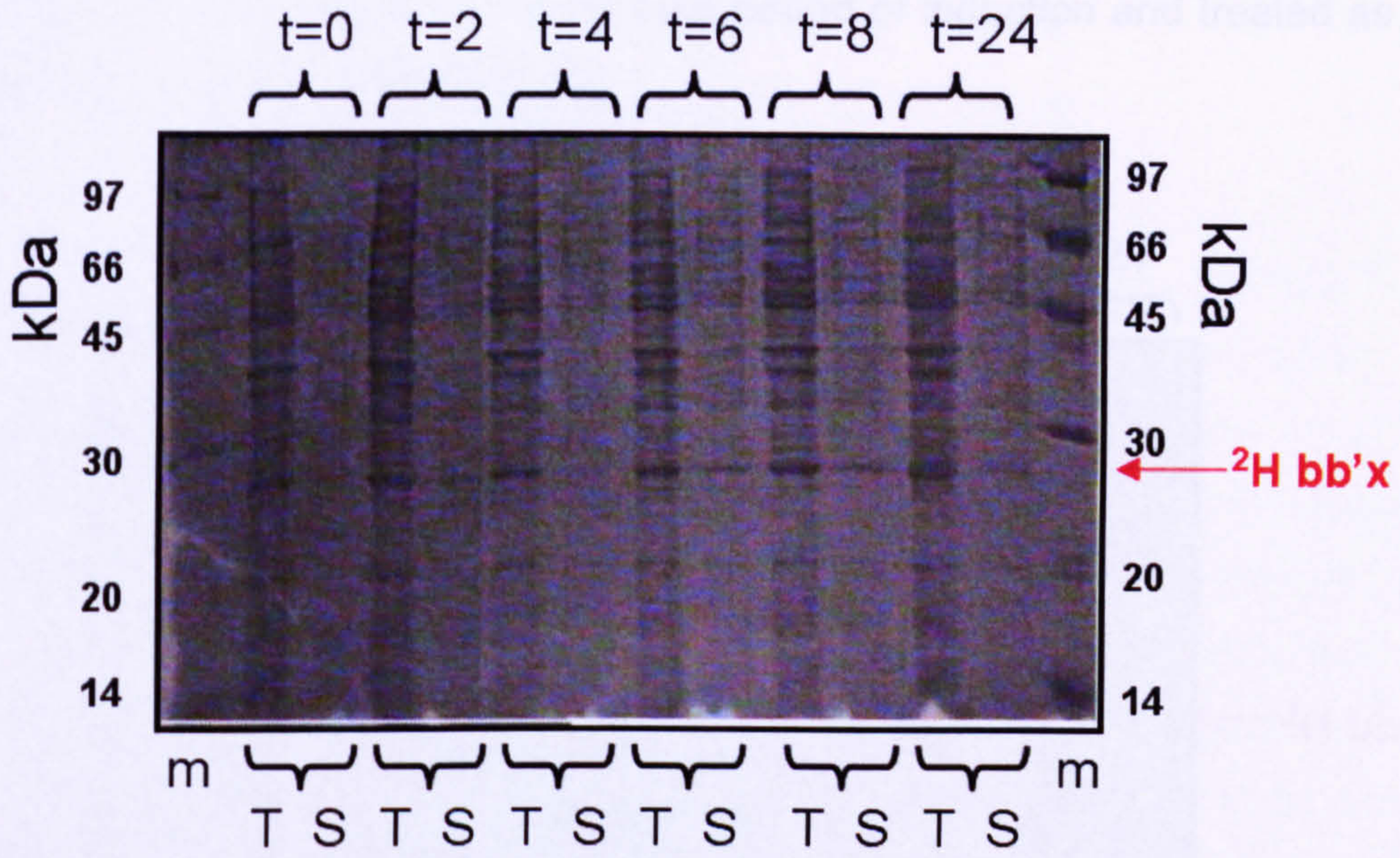

Figure 4-2. 16 % SDS-PAGE of bb'x 96% D_2O test expression. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), m = low molecular weight marker (kDa).

From Figure 4-2 it appears that a very low, if any, expression of soluble **bb'x** is produced with an optimum time of expression around 6 to 8 hours. The 24 hour sample shows a clear reduction of soluble protein, since the band in 24 hour soluble lane is less visible than the soluble protein sample taken at 6 and 8 hours. The total protein expression at 96% D_2O reveals a very low level of expression when compared to SDS-PAGE of $^{13}C/^{15}N$ test expressions in section 3.3.4. Also, all of the ^{2}H soluble protein bands in Figure 4-2 are very small; indicative of a low level of soluble protein expression. It was therefore attempted to re-adapt the cells to improve protein yields but both attempts led to comparable levels of protein expression as shown in the gel in Figure 4-2 (readapted cell expression data not shown). An alternative method of possibly increasing the protein yield would be to express protein at a lower D_2O concentration; this method is fraught with the risk of also reducing the level of ^{2}H incorporation. The 86% D_2O adapted cells were grown in a 40 mL minimal medium culture containing 86% D_2O from the glycerol stock cultures. Again 1

mL samples were taken over a 24 hour period of induction and treated as done previously and run on SDS-PAGE.

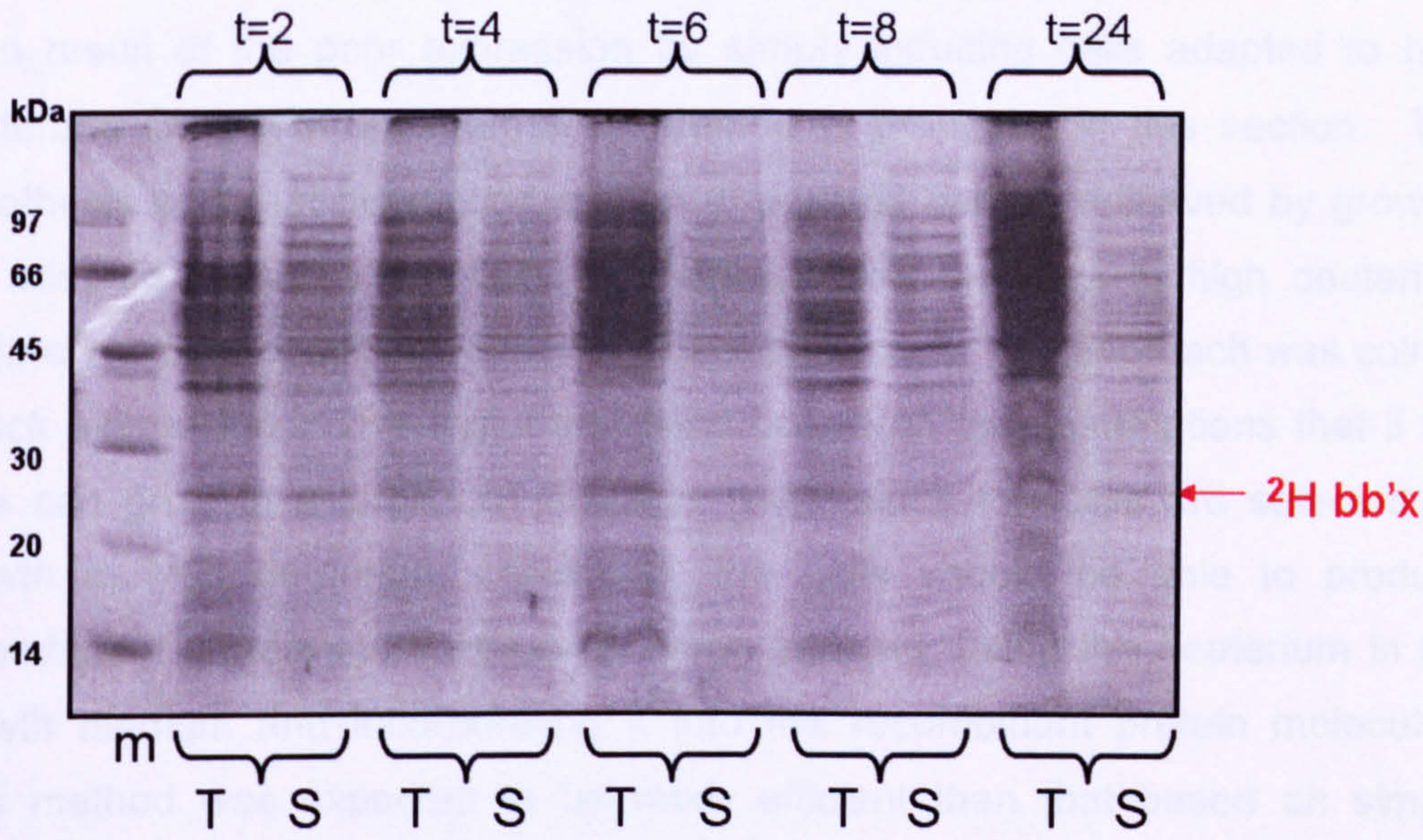

Figure 4-3. 16 % SDS-PAGE of bb'x 86% D_2O test expression. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), m = low molecular weight marker (kDa).

From the expression at 86% D_2O conditions it was clear that there was still a very low amount (if any) of soluble **bb'x**, as the band at approximately 27 kDa, labelled ^{2}H **bb'x** in Figure 4-3, is no more visible than in the 96% D_2O expression test in Figure 4-2 and is very faint. Since a 300 µL, 1.5 mM sample was required for NMR spectroscopy and expression of these cells would require a very large volume culture as with the 96% D_2O expression culture. Again, this was not desirable as D_2O was very expensive and considering the high amount of ^{13}C-glucose and ^{15}N-ammonium sulphate required the cost would be huge. Also this 86% D_2O would likely have a low isotope incorporation rate. Clearly an alternative method of producing $^{2}H/^{13}C/^{15}N$-labelled soluble **bb'x** is required.

4.3. Expression of ^{2}H Labelled Protein Using Adapted Cells and a Shock Expression Technique

As a result of the poor expression by simply inducing cells adapted to high deuterium conditions an alternative method is proposed in this section. The hypothesis was that high incorporation and yields can be achieved by growing the adapted cells in LB initially, then transferring the cells to high deuterium minimal medium conditions and inducing immediately; this approach was coined 'shock expression'. This hypothesis was based on the assumptions that if the cells can grow to a large cell mass in $^{1}H_2O$, since the cells are selected for growth in high deuterium conditions, the cells should be able to produce recombinant protein more efficiently when induced, using the deuterium in the growth medium and incorporating it into the recombinant protein molecules. This method was expected to be more efficient than that based on simple growth and induction, whereby all cellular pathways could be affected by deuterium incorporation.

4.3.1. Test expression using the cells adapted to 86% and 96% D_20 conditions

The proposed methodology would be to take 96% D_2O adapted cells grown to an O.D. of 0.6 in LB and centrifuged at 15900 x g for 10 min at 4 °C. The cells would then be resuspended in 96% D_2O in minimal medium and immediately induced with 1mM IPTG. Samples were taken and treated as described in Figure 3-6 over 8 hours.

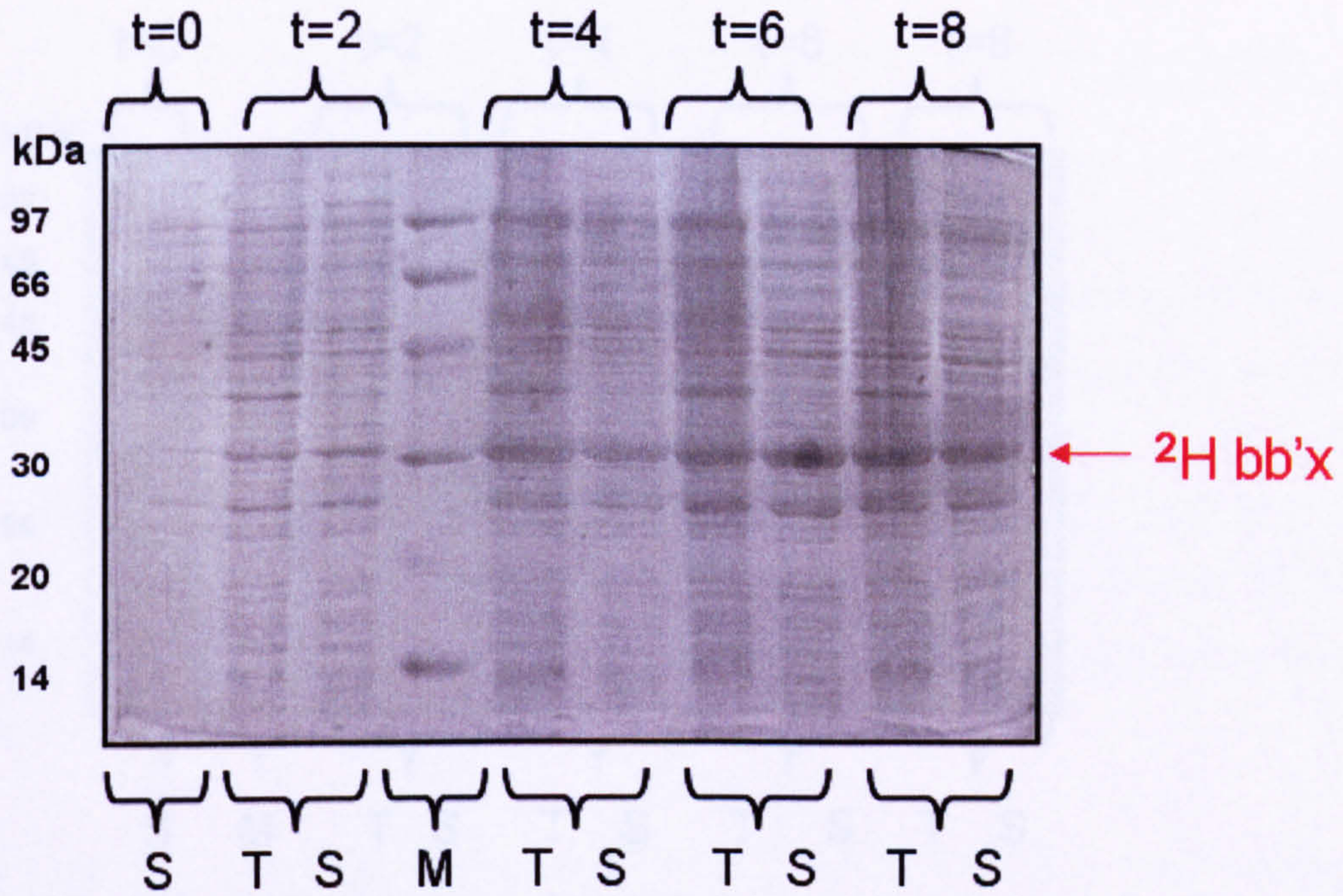

Figure 4-4. 16% SDS-PAGE of bb'x 96% D_2O shock test expression. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), m = low molecular weight marker (kDa).

From the data shown in Figure 4-4 it appears there was induction, since there is no band at 30 kDa in the 0 hour soluble protein sample but there is a band in all the other time point samples at this molecular weight, hence this band is suggestive of labelled soluble **bb'x** expression. The band is far more visible than in Figure 4-2, where the 96% D_2O adapted cells were not shock induced.

Cells adapted to 86% D_2O were also tested; they were resuspended in 86% D_2O to see if a lower D_2O would increase yield and determine if there is an affect on deuterium incorporation. Again samples were taken and treated as described in Figure 3-6 over 8 hours to compare with the 96% D_2O shock induction results.

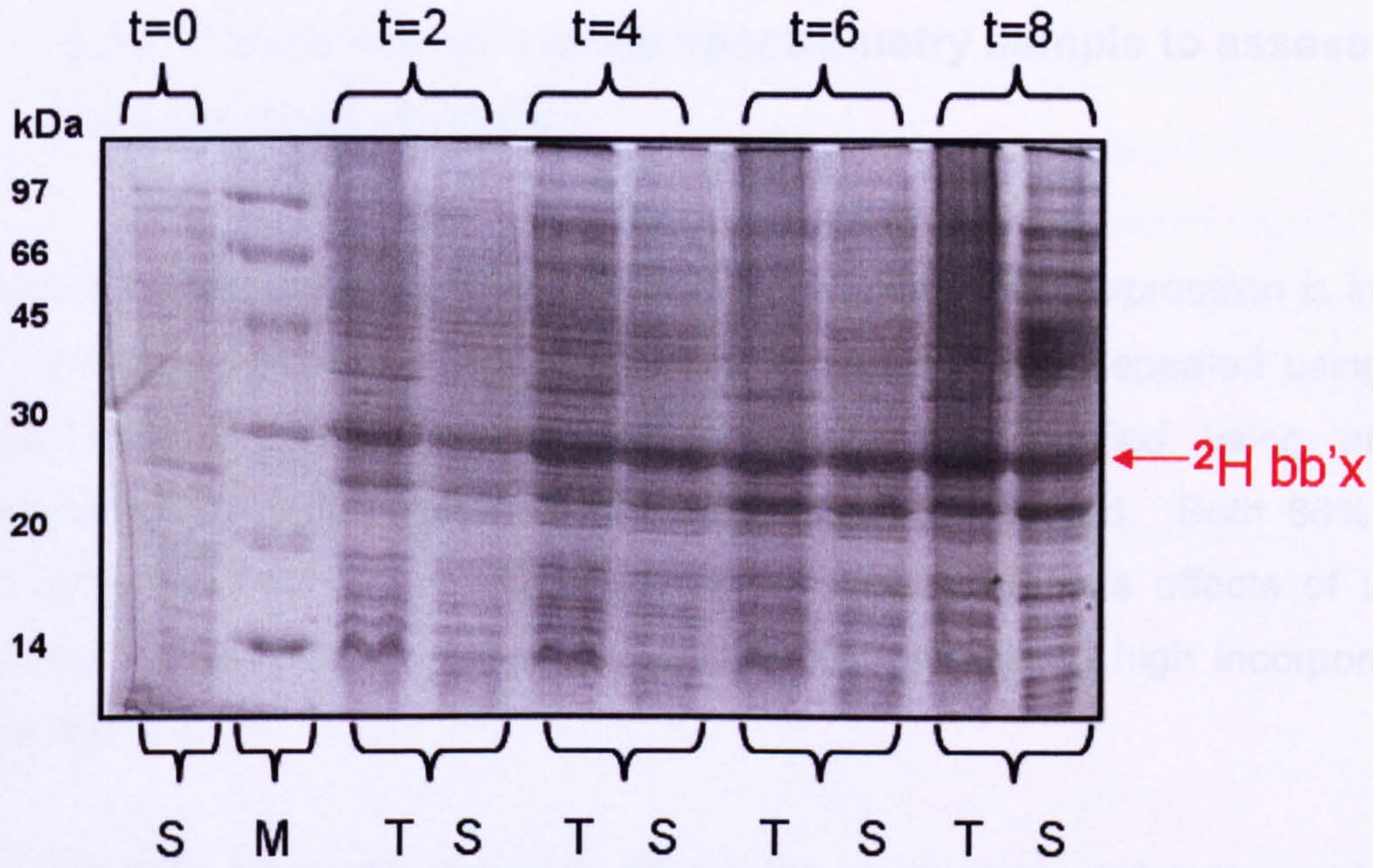

Figure 4-5. 16% SDS-PAGE of bb'x 86% D_2O shock test expression. T = Total protein sample, S = Soluble protein sample, t = time of sample taken after induction (Hours), m = low molecular weight marker (kDa).

It is clear in the 86% D_2O shock induction gel shown in Figure 4-5 that the band at 30 kDa is the induced recombinant protein, since this band is not visible in 0 hour soluble protein sample. Furthermore, as expected the amount of induced protein is greater in this lower 86% D_2O concentration expression than 96%, but it is expected that the incorporation would be lower. But in a positive view, these expression levels appear more promising in their ability to produce a sufficient NMR sample than the previous method of adaptation and induction attempted. These experiments were repeated to confirm reproducibility, which was successful and both sets of experiments produced comparable levels of recombinant protein running at the same molecular weight on the gel (gels not shown).

4.3.2. Production of a mass spectrometry sample to assess incorporation efficiency

To confirm the band around 30 kDa in the preliminary test expression is in fact the ^{2}H labelled soluble recombinant **bb'x**, expression was repeated using the shock induction method, the protein product was purified using affinity chromatography and a mass spectrometry sample prepared. Both 86% and 96% adapted cells were used to identify any advantageous effects of using lower D_2O conditions in relation to yield whilst maintaining high incorporation efficiency.

86% and 96% cells were grown in 40 mL LB, spun down and resuspended in 86% and 96% D_2O conditions in minimal medium, respectively and immediately induced as described in section 4.3.1. The cells were induced for 6 hours as there appears very little difference in yield between 6 and 8 hour of induction shown in the test induction experiments in section 4.3.1. The crude extract was then prepared as described in section 2.2.7 and run on an affinity column.

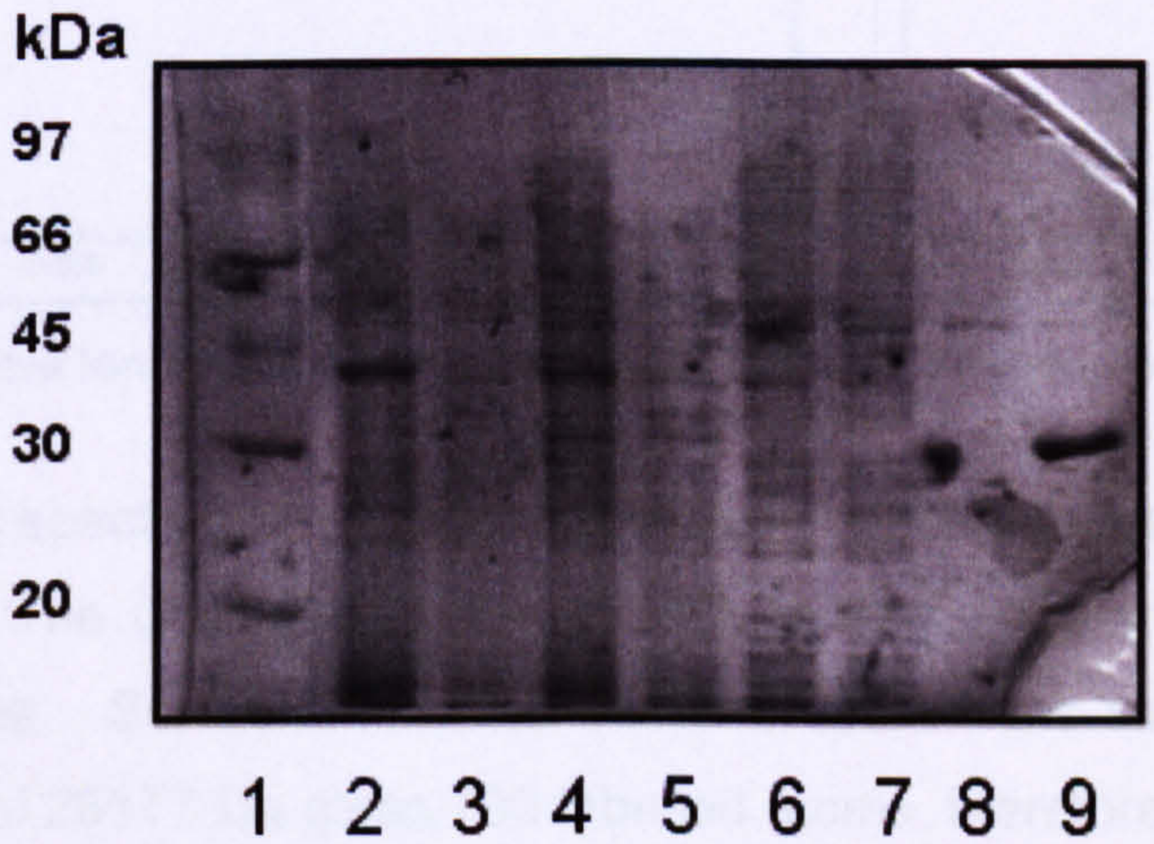

Figure 4-6. 16% SDS-PAGE of 86% D_2O minimal medium expression and IMAC purification. Lane 1, low molecular weight marker. Lane 2, 0 hour expression total protein. Lane 3, 0 hour expression soluble protein. Lane 4, 6 hour expression total protein. Lane 5, 6 hour expression soluble protein. Lane 6, flow through. Lane 7, imidazole wash. Lane 8, low salt wash. Lane 9, hexa-histidine tagged ^{2}H labelled bb'x protein elute.

It is apparent from lanes 6 to 9 in Figure 4-6, of the affinity chromatography purification of 86% shock induced culture, that the purification of a labelled sample was effective, highlighted by the single band in the elute. Since there appears to be very few protein impurities, it was possible therefore to immediately prepare a sample for mass spectrometry analysis; this was done by buffer exchange into 10mM ammonium bicarbonate. The concentration of the elute fraction was measured to be 40 µM.

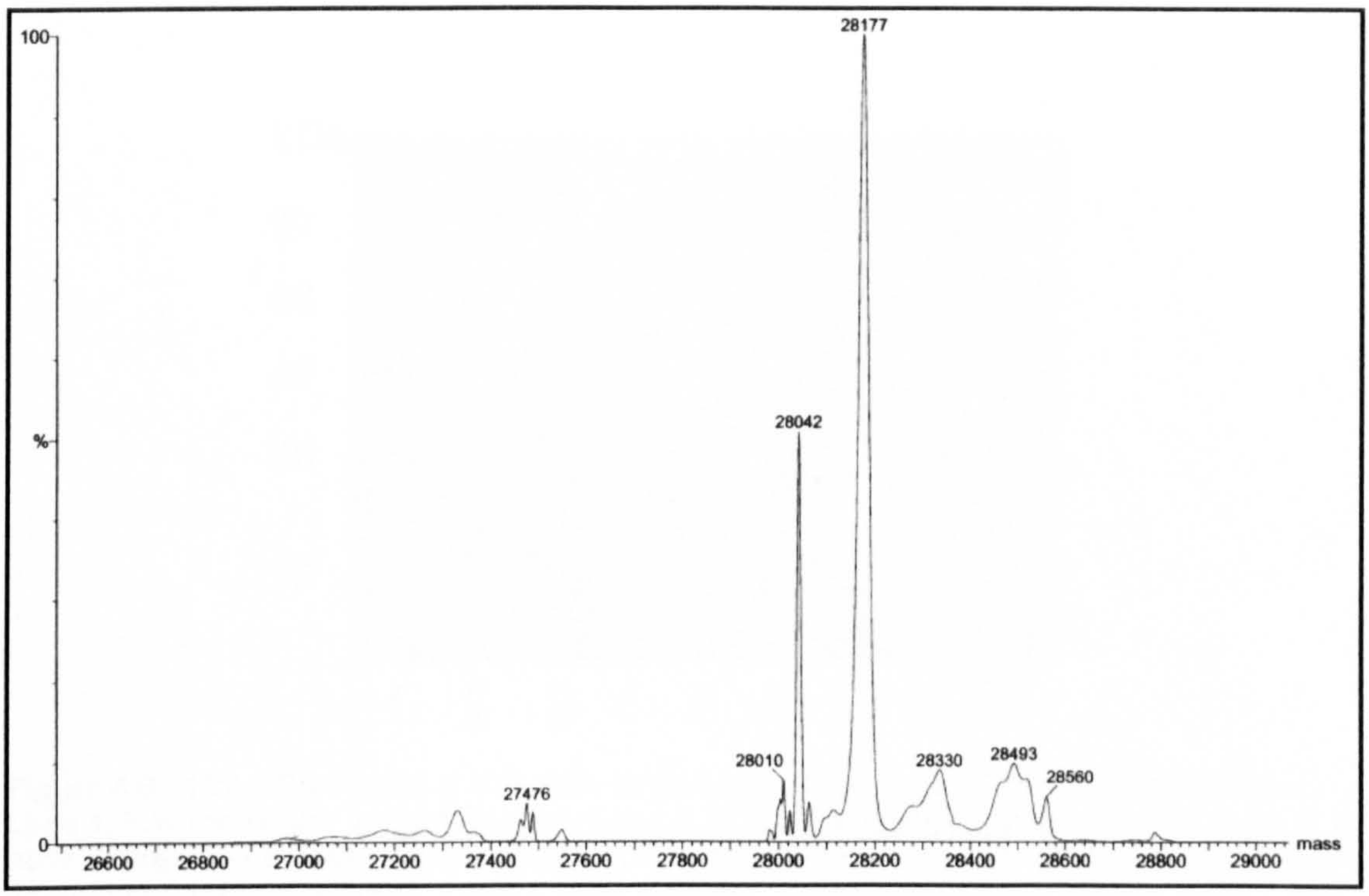

Figure 4-7. Positive ion mass spectrum of 86% D_20 shock induced bb'x.

From the mass spectrometry analysis, the main species has a molecular weight of 28177 Da. The unlabelled mass of **bb'x** is 27477 Da and there are 1910 hydrogen atoms. So the unlabelled mass of 27477 Da subtracted from the labelled mass of 28177 Da gives 700 labelled atoms, therefore (700/1910) x 100 gives a 36.6% ^{2}H labelling efficiency.

There are several other proteins species visible in the mass spectrometry analysis; a peak at 27476 Da shows a very small amount of completely

unlabelled material. There is also peak, at half the intensity of the main peak at 28042 Da, which is minus 135 Da; it was unknown what this species is. It is highly unlikely that this species results from reduced isotope incorporation. Several other unknown minor species also exist at 28330, 28493 and 28560 Da.

Perhaps the lower amount of D_2O in the expression medium might be the cause of the low deuterium incorporation; hence the 96% D_2O minimal medium expression can be compared to test this hypothesis.

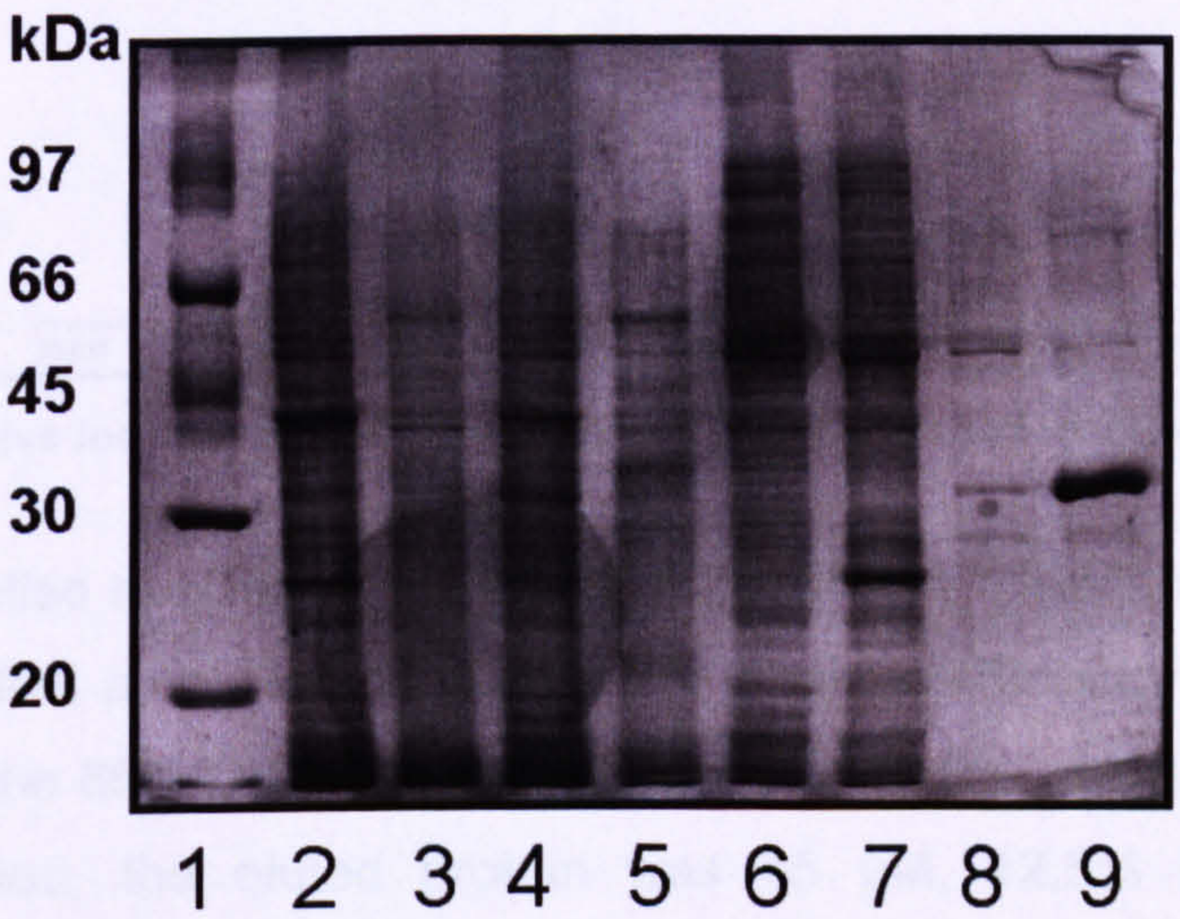

Figure 4-8. 16% SDS-PAGE of 96% D_2O minimal medium expression IMAC purification. Lane 1, low molecular weight marker. Lane 2, 0 hour expression total protein. Lane 3, 0 hour expression soluble protein. Lane 4, 6 hour expression total protein. Lane 5, 6 hour expression soluble protein. Lane 6, flow through. Lane 7, imidazole wash. Lane 8, low salt wash. Lane 9, hexa-histidine tagged ^{2}H labelled bb'x protein elute.

Again the affinity chromatography purification of 96% shock induced culture, was successful in purifying the labelled protein as shown in lanes 6 to 9 in Figure 4-8. The eluted protein has a few impurity bands but relatively clean enough to allow a mass spectrometry sample to be produced. The eluted proteins appear comparable by visually comparing the size of the elute bands in 86 and 96% D_2O conditions.

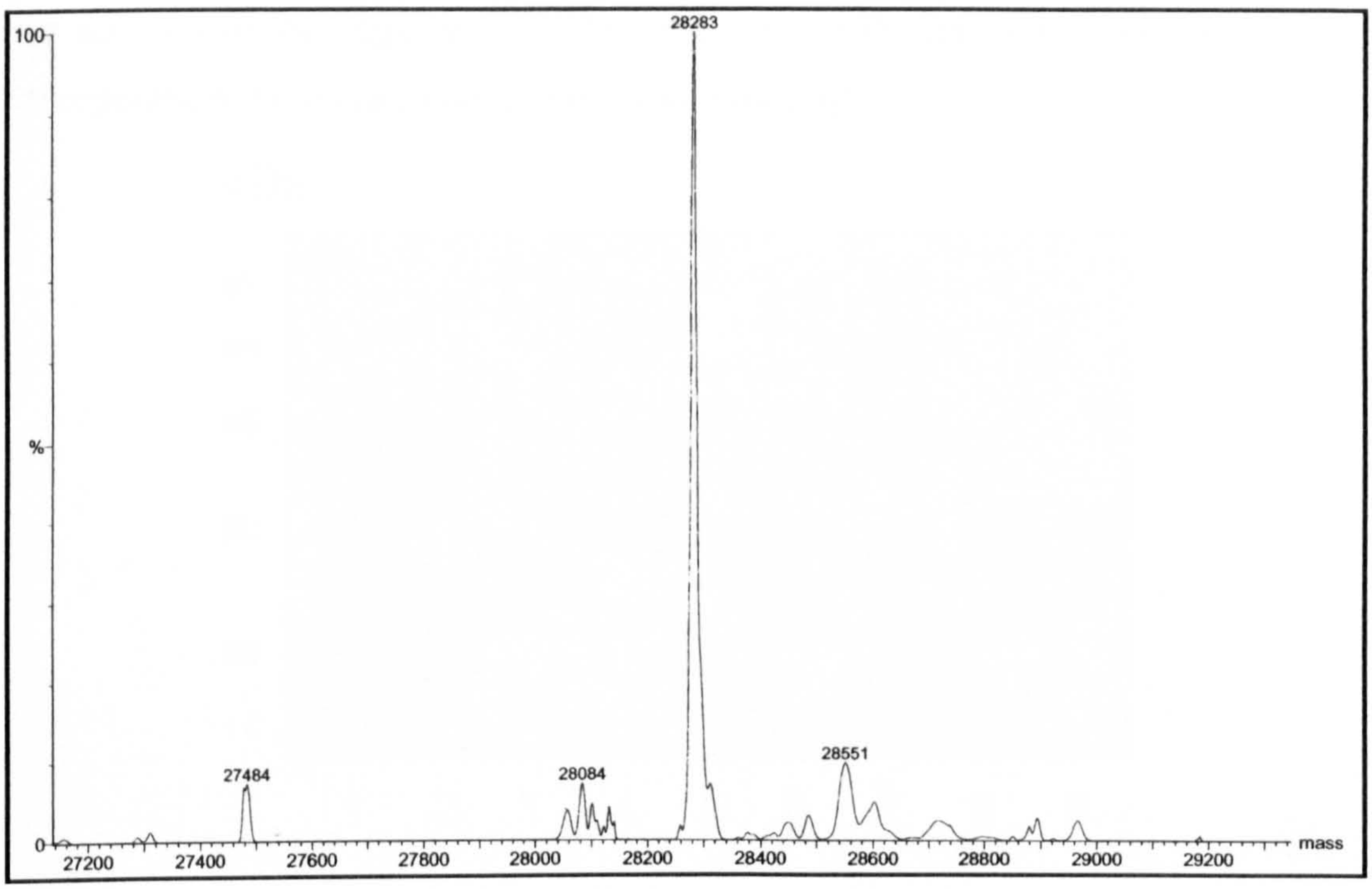

Figure 4-9. Positive ion mass spectrum of 96% D_20 shock induced bb'x.

The major labelled species has a mass of 28283 Da, therefore 28283 – 27477 gives 806 labelled atoms which is a 42.2% labelling efficiency. This is a 5.6% increase from the 86% D_2O labelling efficiency of 36.6%. Using the 96% shock induction method, the eluted protein was 45 μM, 12.5% increase to that produced in the 86% D_2O in minimal medium. If these experiments were scaled up 20 fold to the typical 800mL expression then theoretically a 0.8 mM sample should be produced; this concentration was too low for a NMR sample. Hence it was most important to increase the labelling efficiency and some improvement of yield.

4.3.3. Test expression using un-adapted cells and shock expression to test the necessity of adaptation

To determine whether adaptation of cell was required, cells un-adapted to high deuterium conditions using the process described in section 4.2 are grown using

the same methodology as with 86 and 96% adapted cells and the level of incorporation measured using mass spectrometry.

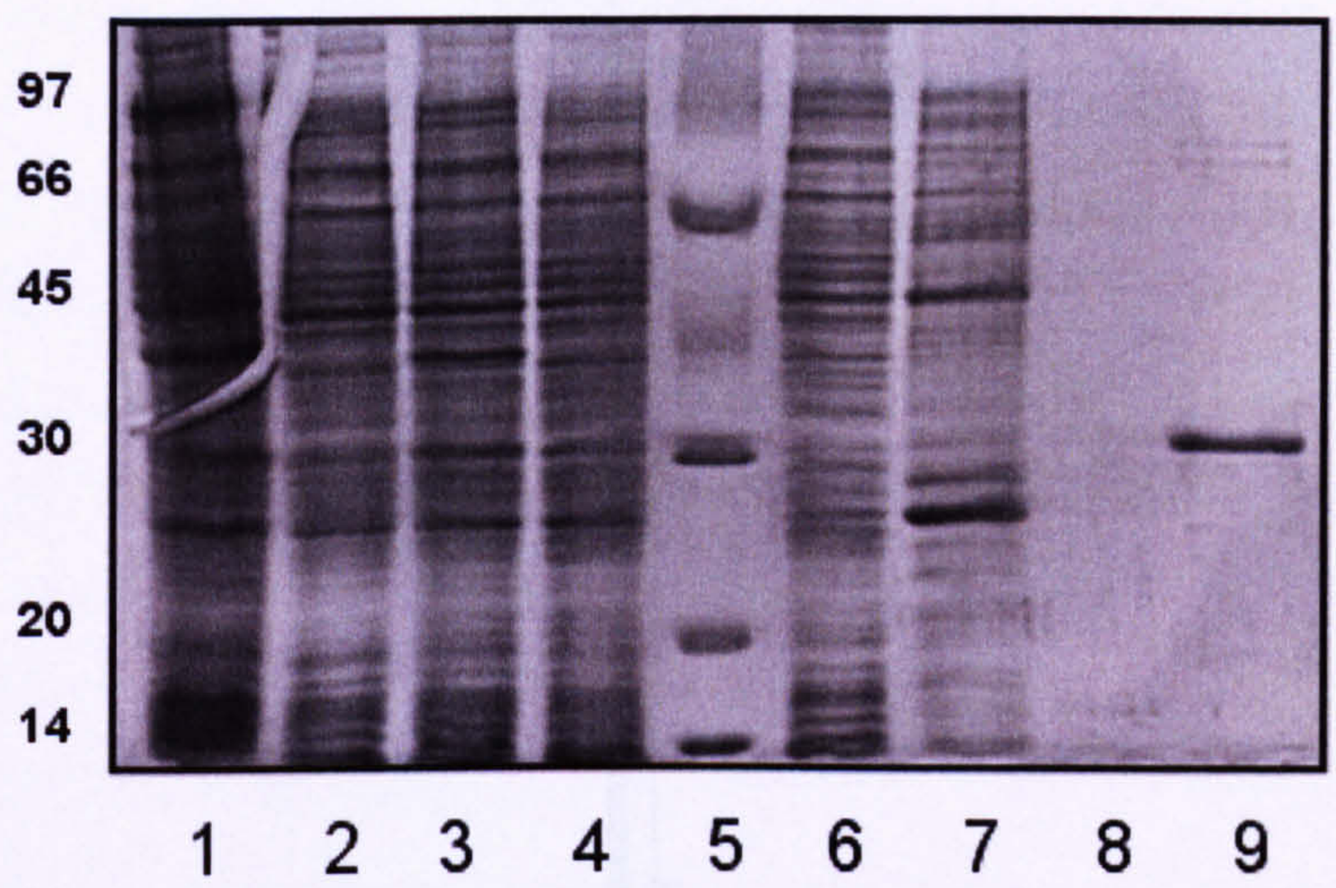

Figure 4-10. 16% SDS-PAGE of the affinity column protein from un-adapted cell shock induction in 96% D_2O. Lane 1, 0 hour induction total protein. Lane 2, 0 hour induction soluble protein. Lane 3, 6 hour induction total protein. Lane 4, 6 hour induction soluble protein. Lane 5, low molecular weight marker. Lane 6, flow through. Lane 7, imidazole wash. Lane 8, low salt wash. Lane 9, hexa-histidine tagged protein elute.

From the gel of unadapted cells shock induced, it was clear the level of expression level was lower than for adapted cell if lane 9 of the soluble protein after IMAC purification and the corresponding bands in the 86% adapted cells Figure 4-6, lane 9 or 96% adapted cells Figure 4-8, lane 9 are compared. This was further verified as the calculated concentration of the IMAC elute after the affinity column was 27 μM, whereas with the adapted cells of 86% and 96% was 40 and 45 μM respectively. From the gel, it was clear that there are very few protein impurities in the eluted protein fraction; the main species believed to be the isotopically labelled recombinant protein appears at a molecular weight of approximately 31 KDa. Since the elute was sufficiently purified a sample was submitted for mass spectrometry analysis.

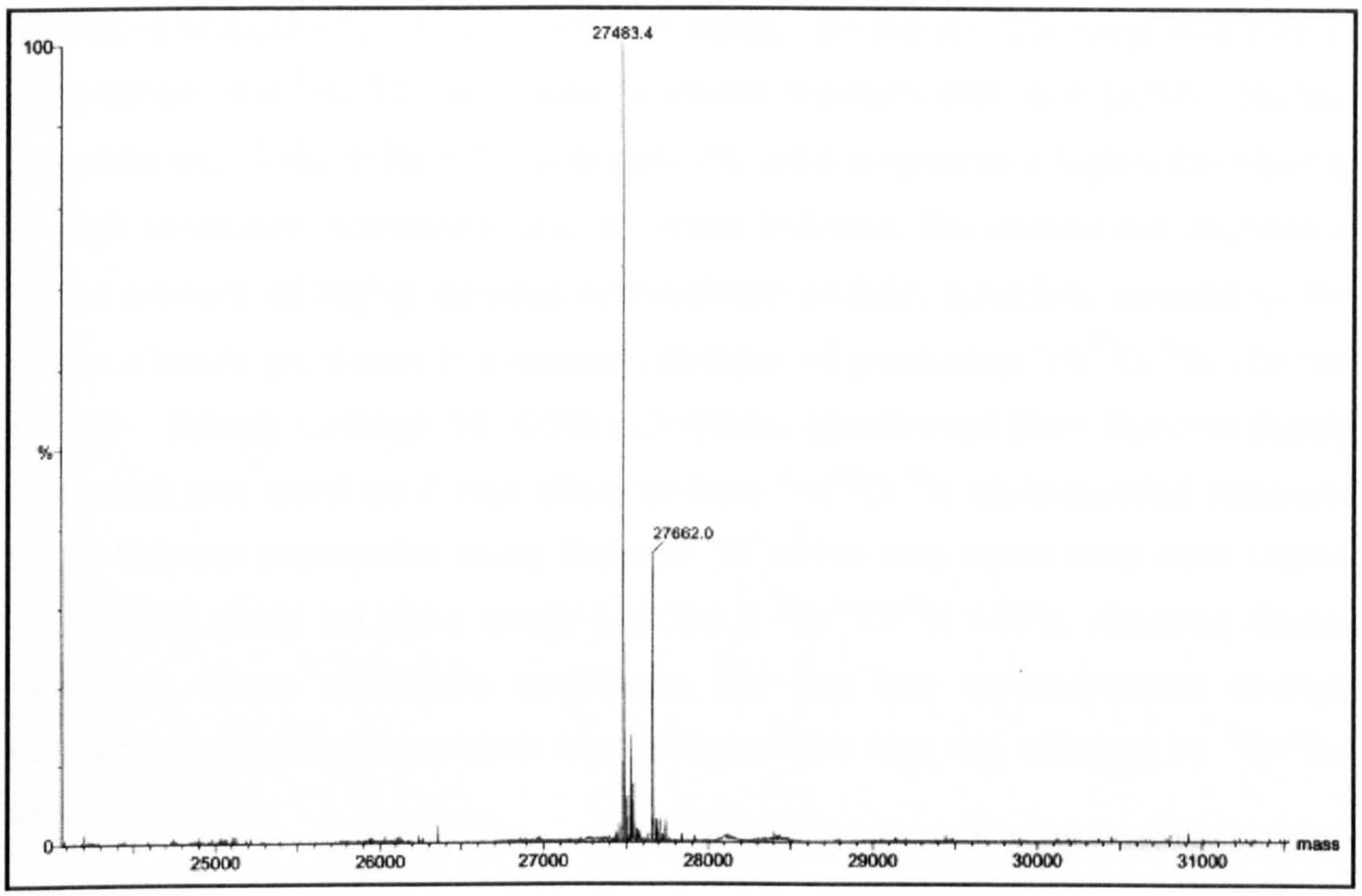

Figure 4-11. Positive ion mass spectrum of the affinity column eluted protein from un-adapted cell shock induction in 96% D_2O.

The main species has a mass of 27483 Da which indicates 6 of the 1910 hydrogen atoms have been labelled, which is a 0.3% labelling efficiency. The low efficiency is likely to be a result of the cells not being adapted to high D_2O conditions. Therefore it was imperative that any further work done to improve yield and more importantly labelling efficiency must be carried out using the adapted cells. The other minor species at 27662 Da is an extra 179 which results from the dephosphorylated form of the α-N-6-Phosphogluconoylation of the histidine tag.

4.4. Use of Isotopically Labelled Rich Medium with Adapted Cells to Increase Isotope Incorporation Efficiency

There are a host of isotopically labelled rich media commercially available, but the draw back was the high price for large quantities required to produce

sufficient amount of protein for NMR analysis. So the method proposed was to supplement the $^{2}H/^{13}C/^{15}N$ labelled minimal medium with isotopically labelled rich medium. This, in theory, will enable the cells to grow to a higher cell density in high deuterium conditions and so, once induced, the culture will express a larger amount of highly labelled recombinant protein; hopefully exceeding the levels already produced in previous attempts of producing $^{2}H/^{13}C/^{15}N$ labelled protein. Initially Celtone ^{2}H >97% rich media (purchased from Spectra Stable Isotopes) was used as it was cheaper than $^{2}H/^{13}C/^{15}N$ triple-labelled medium, but if the test expression using Celtone ^{2}H >97% was successful then further expression could be done using Spectra 9 $^{2}H/^{13}C/^{15}N$ >97% (Spectra Stable Isotopes), since deuterium conditions are the key environmental change believed to be affecting protein expression which was not affected by $^{13}C/^{15}N$-labelling.

4.4.1. Test expression using 5% and 10% Celtone ^{2}H rich medium

The 96% adapted cells, were re-grown from the glycerol stock by streaking on LB agar plates with appropriate antibiotics. A colony from this plate was used to inoculate 40 mL cultures. Two cultures were inoculated, one with 96% D_2O minimal medium and the second with 96% D_2O minimal medium supplemented with 5% (2 mL) of Celtone ^{2}H >97% rich medium. Both were grown and O.D. measured at 600nm over a 24 hour period.

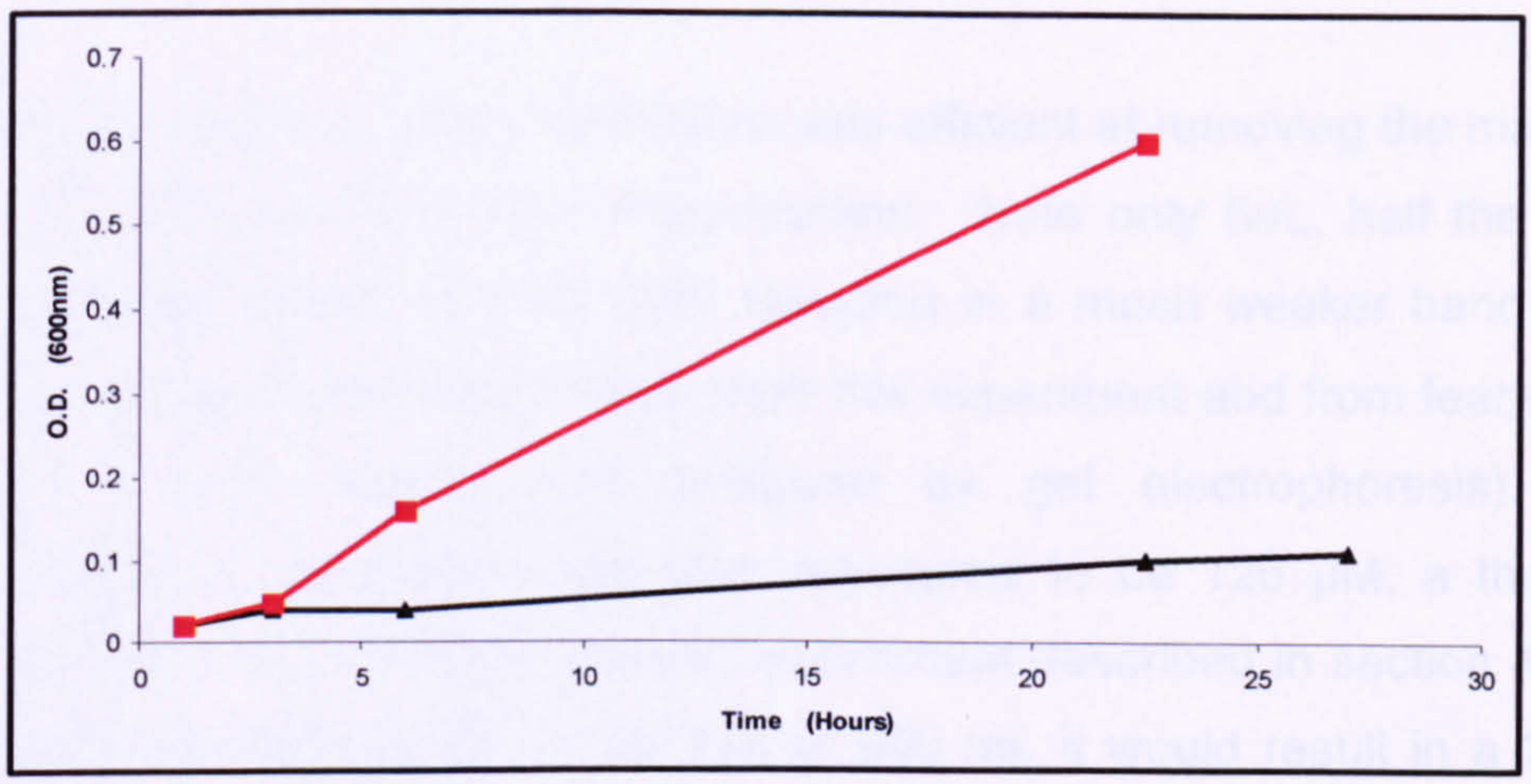

Figure 4-12. Growth curves of cultures grown in 96% D_2O minimal medium (black line) and D_2O 96 % minimal medium supplemented with Celtone ^{2}H (red line).

By comparing the relative cell densities, it was clear that the Celtone ^{2}H medium leads to a much higher growth rate and cell density. The Celtone ^{2}H medium expression was then induced with IPTG at the 23 hour time point for 6 hours and products purified using affinity chromatography.

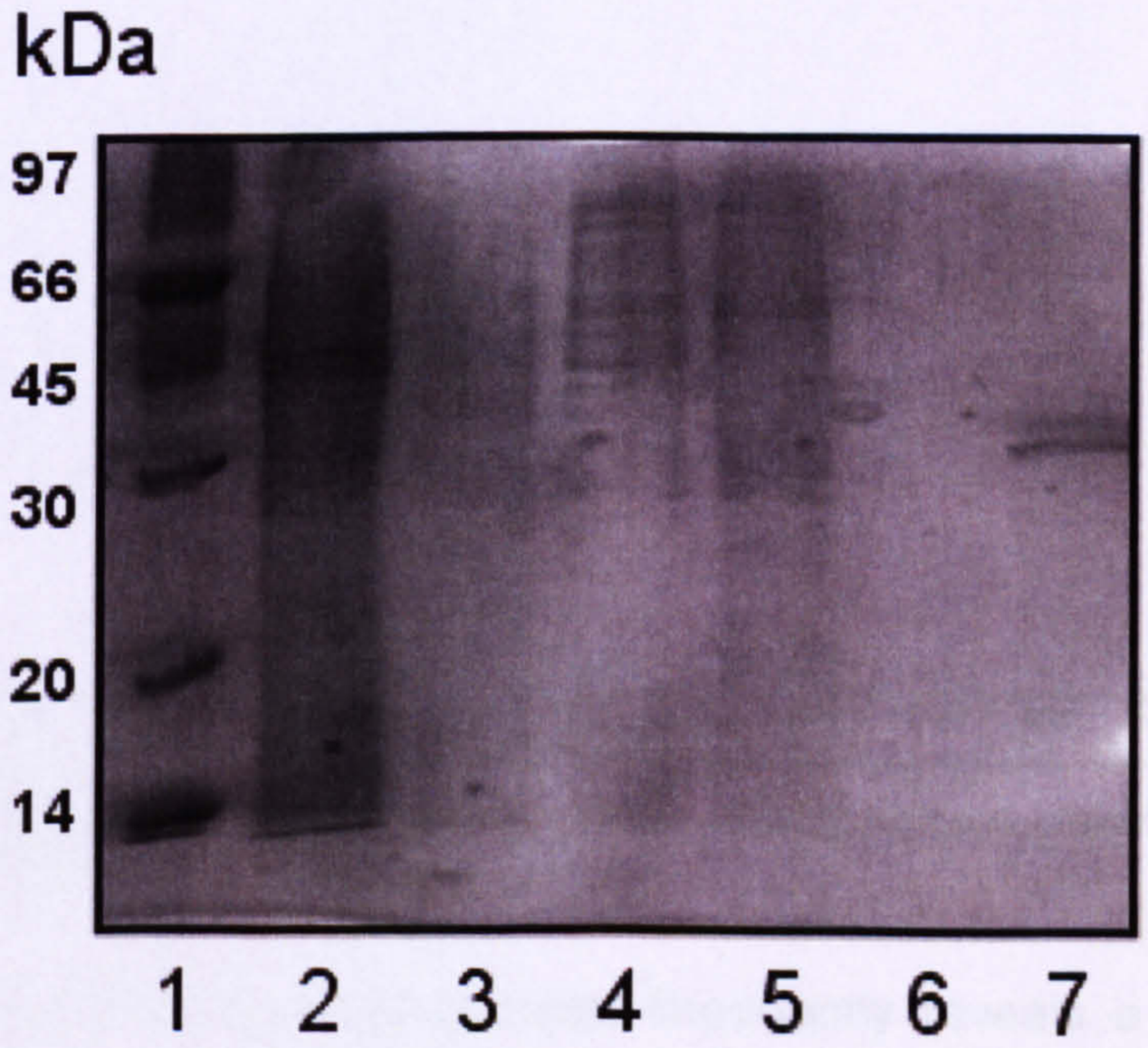

Figure 4-13. 5% Celtone ^{2}H 40 mL test expression. Lane 1, low molecular weight marker. Lane 2, total protein. Lane 3, soluble protein. Lane 4, flow through. Lane 5, imidazole wash. Lane 6, low salt wash. Lane 7, hexa-histidine tagged bb'x protein elute.

As shown in Figure 4-13 the purification was efficient at removing the majority of proteins, leaving a highly pure elute fraction. Note only 5uL, half the normal volume usually loaded into the gels, resulting in a much weaker band (it was expected a higher yield would result from this experiment and from fear of over-loading a lower volume was analysed by gel electrophoresis). The concentration of the protein was then measured to be 126 μM, a three fold increase from the 96% shock induced experiment described in section 4.3.2. If this experiment was scaled up 20 fold to 800 mL it would result in a 2.5 mM sample, which is borderline sufficient to produce a NMR sample; if you consider 20-30% losses during purification and preparation. The sample was then prepared for mass spectrometry to measure the level of isotope incorporation.

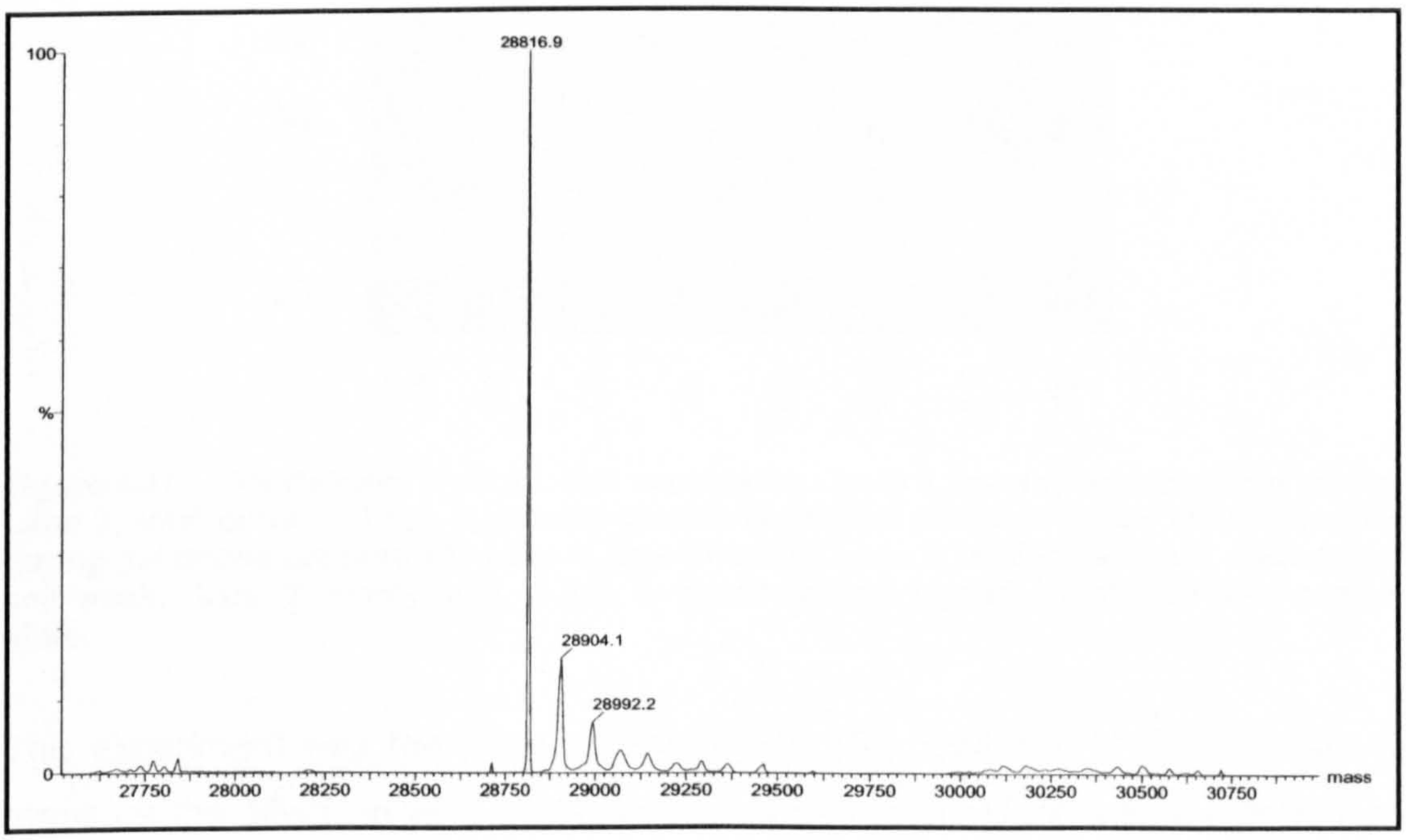

Figure 4-14. Positive ion mass spectrum of the affinity column eluted protein from 5% celtone ^{2}H supplemented medium.

The mass spectrometry analysis more importantly reveals a major peak at 28817 Da, which has 1340 labelled isotope out of a possible 1910 possible atoms and so is a 70.2% labelling efficiency. This is a 28% increase from the 96% shock induction expression in section 4.3.2. It may be possible to increase

protein yield by supplementing with more Celtone rich medium. Therefore the next proposed experiment was to repeat the expression with 10% Celtone ^{2}H rich medium and the eluted protein concentrations compared.

The other minor species has a 175 Da larger mass than the major species of 28817 Da and so is likely to be the dephosphorylated α-N-6-Phosphogluconoylation of the histidine tag. A very minor unknown species was also seen at 28904 Da.

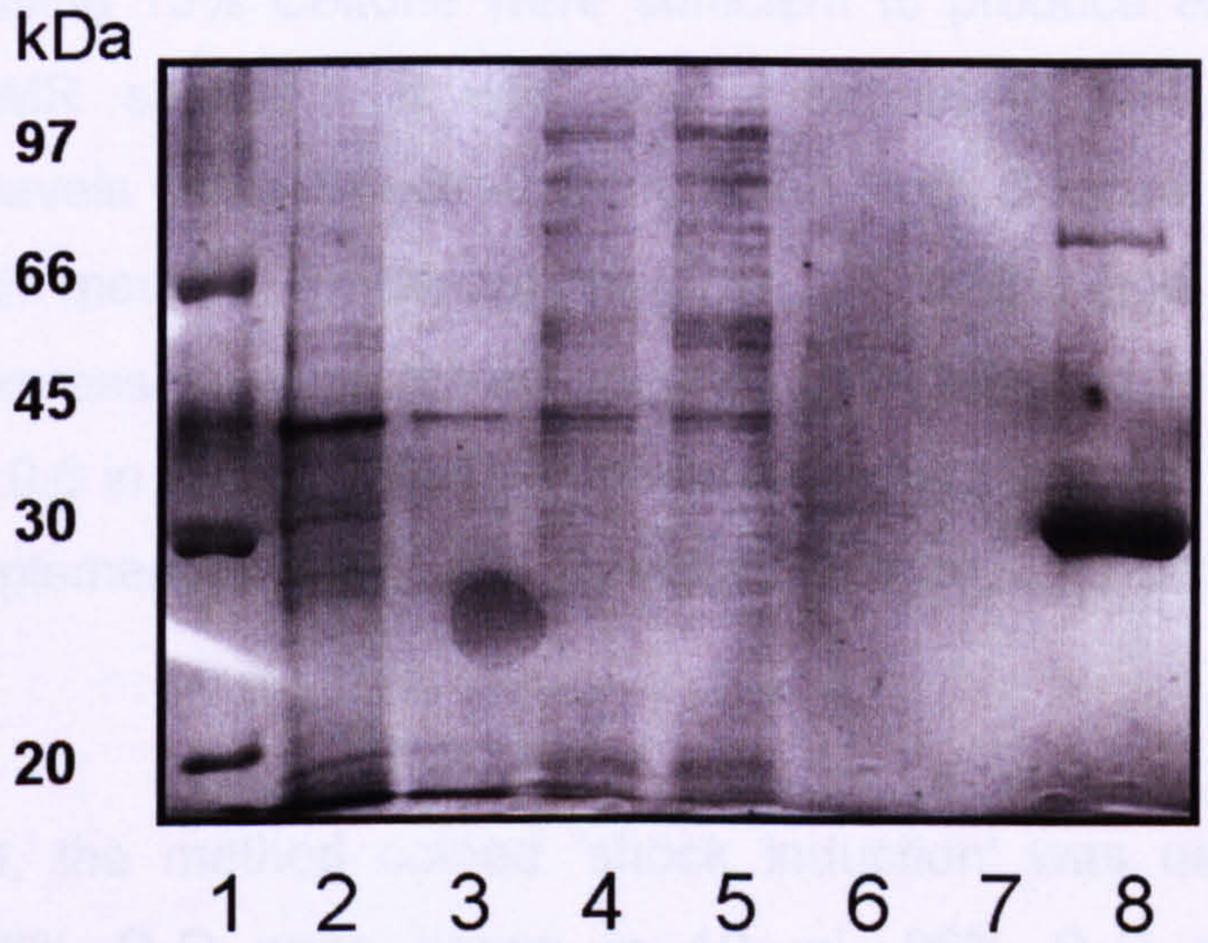

Figure 4-15. 10% Celtone ^{2}H 40 mL test expression. Lane 1, low molecular weight marker. Lane 2, total protein. Lane 3, soluble protein (spherical shape is an air bubble trapped during gel drying procedure). Lane 4, flow through. Lane 5, imidazole wash. Lane 6, low salt wash. Lane 7, empty lane. Lane 8, hexa-histidine tagged ^{2}H-labelled bb'x protein elute.

The experiment was therefore repeated with 10% Celtone ^{2}H rich medium to measure the affect on yield and isotope incorporation. Upon visual comparison of the two elute fractions in Figure 4-13 and Figure 4-15, the elute fraction in the 10% Celtone is much bigger (even though double the volume was loaded here). When the protein concentration was measured, it revealed an increase of 37% from 126 μM (5% celtone) to 200μM (10% celtone). Therefore if the 10% Celtone experiment was scaled up 20 fold to 800 mL then an elute sample of 4 mM would produced, which with losses during purification would be

sufficient to produce a NMR sample. The isotope incorporation was not measured by mass spectrometry analysis, but since the incorporation measured using 5% Celtone was 70.2%, with 10% Celtone the level of incorporation can be expected to be comparable if not higher.

4.4.2. Test expression using cells adapted to 96% D_2O grown in Spectra9 rich medium and shock induced

Experiments using 10% Celtone were sufficient to produce enough protein to produce a NMR sample. It was now essential to test expression and incorporation levels using Spectra9 (purchased from Spectra Stable Isotopes) which is a rich medium containing ^{2}H>97%, ^{13}C >98% and ^{15}N>98%. Two methods of expression were tested; in one, cells adapted to 96% D_2O were grown to O.D. 0.6 in 40 mL 96% D_2O minimal medium with ^{13}C and ^{15}N-labelled reagents, supplemented with 10% Spectra9 culture, induced and harvested after 6 hours.

In the second, the method coined 'shock induction' was used, where cells adapted to 96% D_2O were grown in 40 mL 96% D_2O minimal medium supplemented with 10% Spectra9 rich medium to O.D. 0.6, spun down and re-suspended in 40 mL 96% D_2O in minimal medium (at O.D. 0.1) with ^{13}C and ^{15}N-labelled reagents, grown to O.D. 0.6, induced and harvested after 6 hours (note no rich medium in expression culture). The proteins produced by both methods were then purified by affinity chromatography.

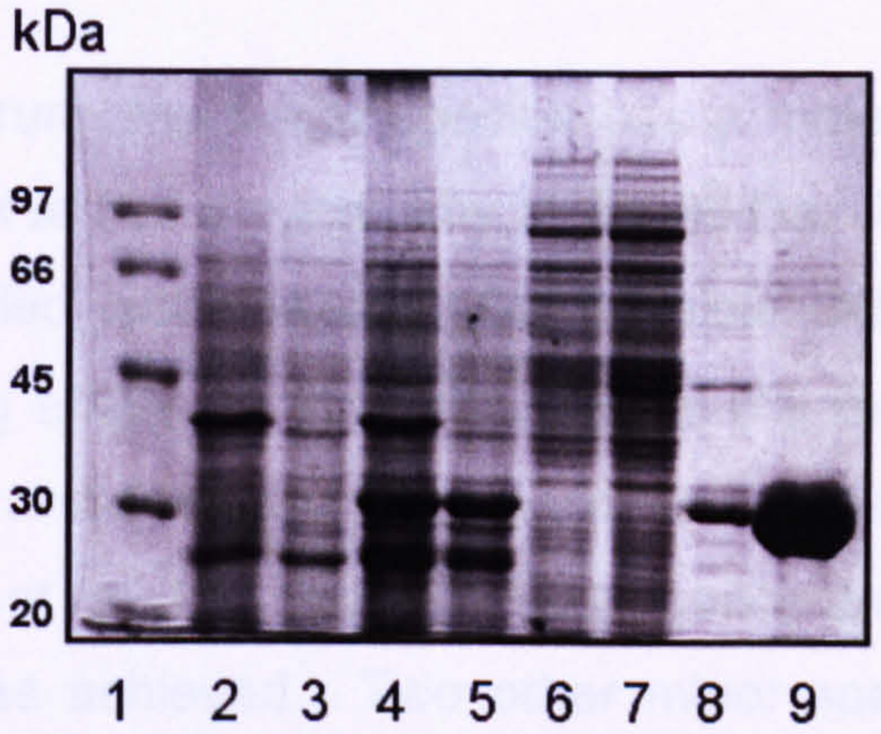

Figure 4-16. 10% Spectra9 40 mL simple induction test expression. Lane 1, low molecular weight marker. Lane 2, total protein before induction. Lane 3, soluble protein before induction. Lane 4, total protein after 6 hour induction. Lane 5, soluble protein after 6 hour induction. Lane 6, flow through. Lane 7, imidazole wash. Lane 8, low salt wash. Lane 9, hexa-histidine tagged triple labelled bb'x protein elute.

From Figure 4-16, it is clear that expression levels are much higher than seen previously, as the size of the elute band is very large. The concentration of the elute fraction was measured to be 0.5mM. Also the purification using affinity chromatography was successful at removing the majority of protein impurities and so the elute was prepared for mass spectrometry analysis.

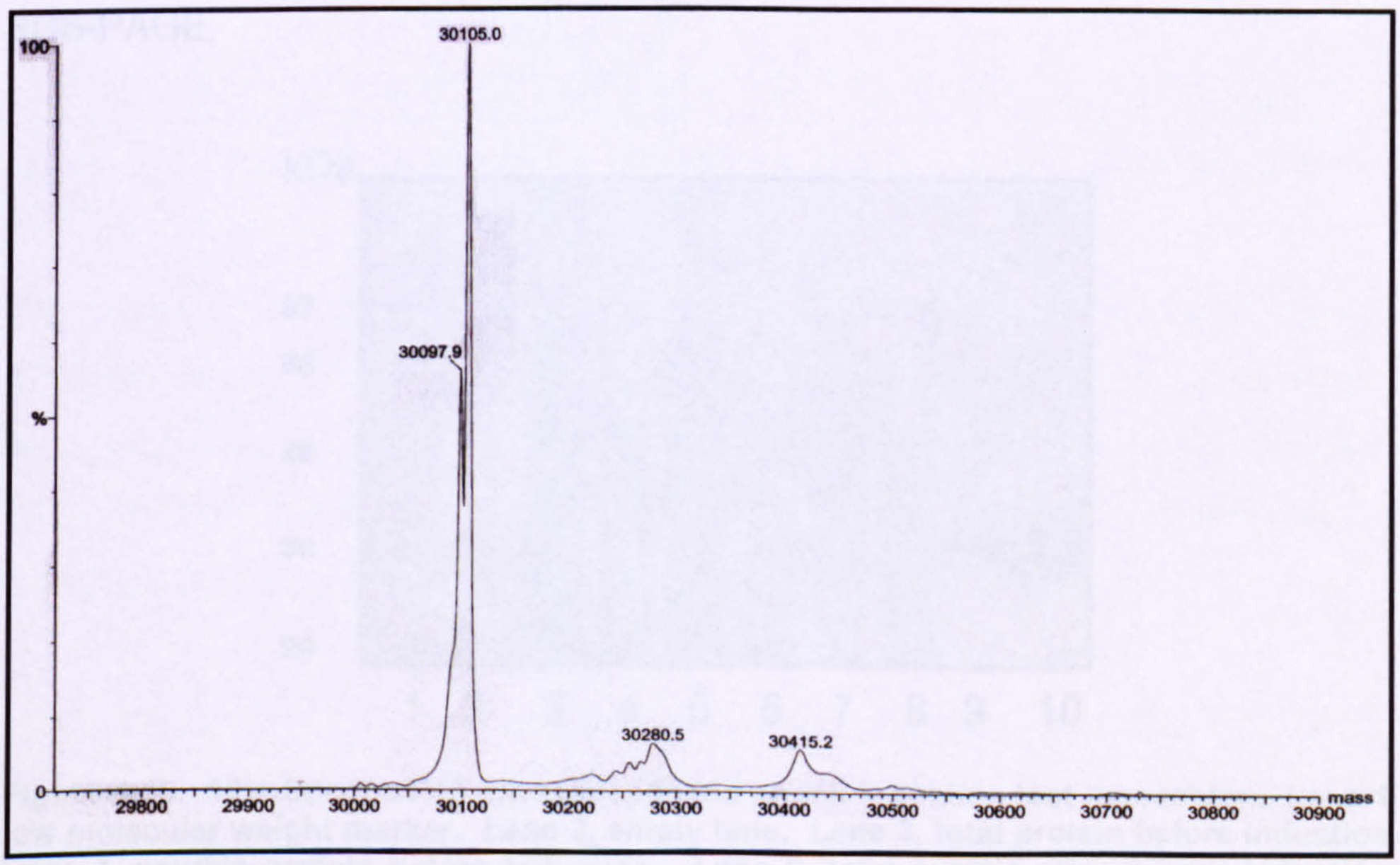

Figure 4-17. Positive ion mass spectrum of 10% Spectra9 40 mL simple induction test expression.

From the mass spectrum, the major species has a molecular weight of 30105 Da and a second peak at half the intensity at 30098 Da. These masses indicate 2628 and 2621 labelled isotopes out of a possible 3468 $^{2}H/ ^{13}C/^{15}N$ atoms, hence a total labelling efficiency of 75.8% and 75.6% respectively. This is an improvement of 5.6% and 5.4% from the previous expressions using the single ^{2}H labelled Celtone at a 5% supplemented concentration whereby 70.2% labelling efficiency was achieved. Two other minor species can be seen by mass spectrometry at 30280 Da which is 175 larger than 30105 Da species and produced by the dephosphorylated α-N-6-Phosphogluconoylation of the histidine tag.

The shock induction method was also tested to see if yields and incorporation efficiency can be increased. Cells adapted to 96% D_2O were grown in a 40 mL 96% D_2O minimal medium supplemented with 10% Spectra9 culture, grown to O.D. 0.6, spun down, resuspended in 96% D_2O, (at O.D. 0.1), grown to O.D. 0.6, induced and harvested after 6 hours. 1mL samples were taken and run on SDS-PAGE.

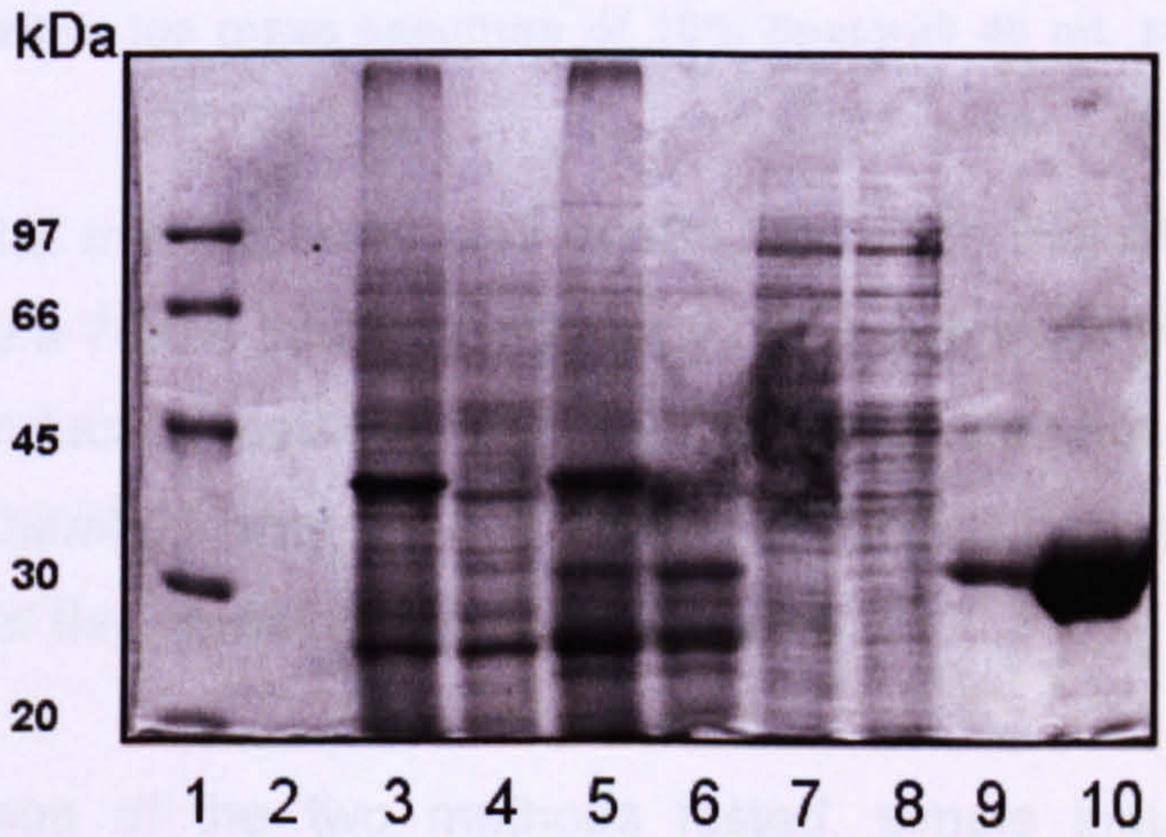

Figure 4-18. 10% Spectra9 40 mL resuspended shock induction test expression. Lane 1, low molecular weight marker. Lane 2, empty lane. Lane 3, total protein before induction. Lane 4, soluble protein before induction. Lane 5, total protein after 6 hour induction. Lane 6, soluble protein after 6 hour induction. Lane 7, flow through. Lane 8, imidazole wash. Lane 9, low salt wash. Lane 10, hexa-histidine tagged bb'x protein elute.

Again the expression appears to be very successful and large protein band is produced as seen in the elute fraction after affinity chromatography purification. The concentration was measured to be 0.5 mM, comparable to the Spectra9 simple induction experiment done previously. The elute was then prepared for mass spectrometry analysis.

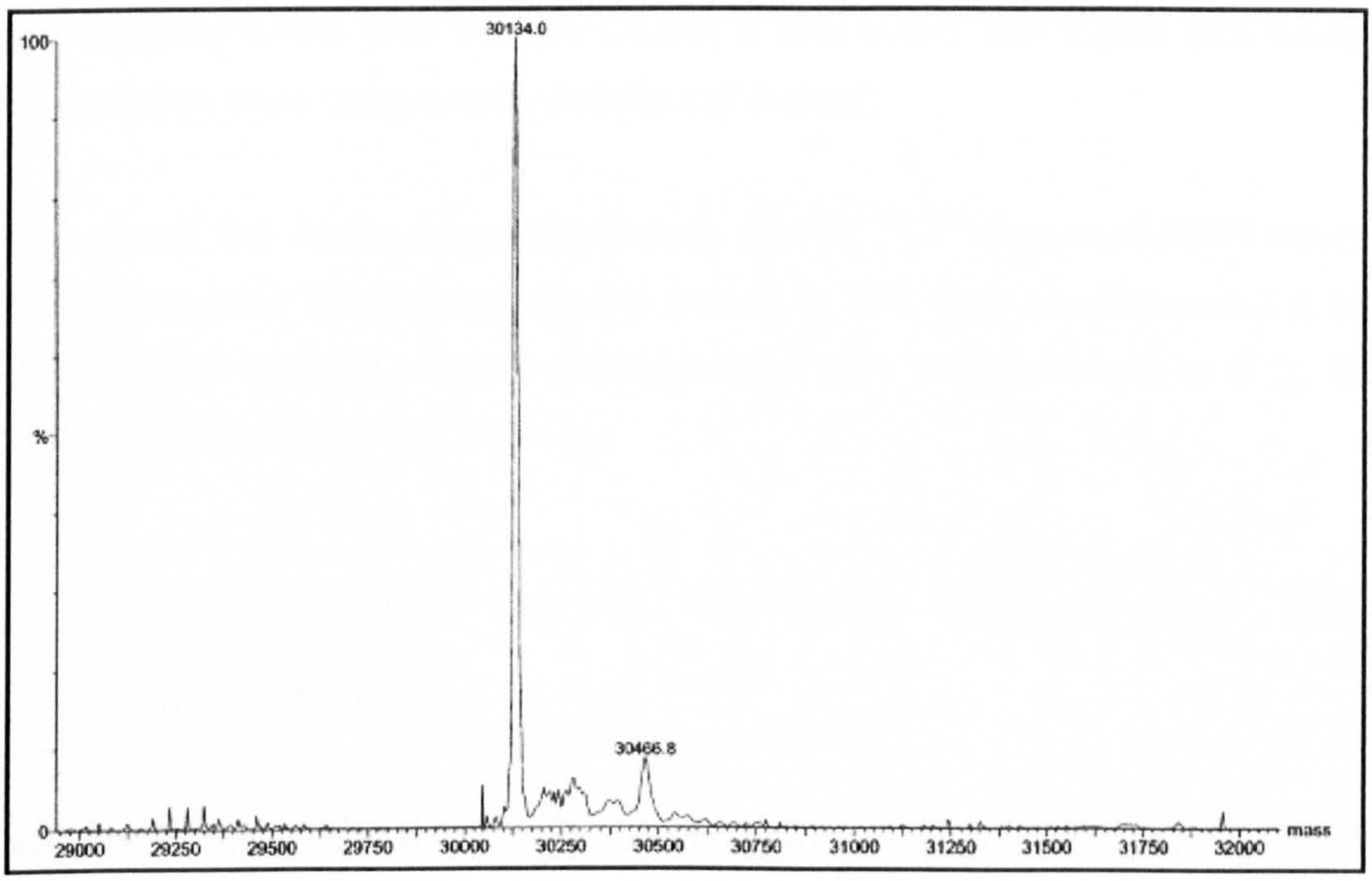

Figure 4-19. Positive ion mass spectrum of 10% Spectra9 40 mL shock induction test expression.

The mass of the main species was 30134 Da which indicates 2657 labelled isotopes, hence a 76.6% labelling efficiency. This is a mere 0.8% improvement on the simple induction method. The other minor species has a mass of 30467 Da, with 2990 labelled isotopes and so has an 86% labelling efficiency but is less than 10% of the intensity of the main species.

Upon comparison of the two methods tested, simple induction and shock induction revealed comparable yields of 0.5 mM which is more than sufficient when scaled up to produce a NMR sample, since scaling up 20 fold to 800 mL should produce a sample of 10 mM more than sufficient for a NMR sample.

Also, very similar labelling efficiencies were recorded of 75.8% and 76.6% respectively, which is sufficiently high enough to carry out the NMR experiments required with a $^{2}H/^{13}C/^{15}N$ sample. Hence in the interest of efficiency, the simple induction is less time intensive, so was preferable. This was the method selected to be employed when scaled up to produce a triple labelled NMR sample. The simple induction experiment was repeated to confirm reproducibility which was achieved here; it was found that yields and isotope incorporation were comparable (results not shown).

So to recap, the methodology to produce the $^{2}H/^{13}C/^{15}N$-labelled NMR sample would be to grow the cells adapted to growth in 96% D_2O conditions in 2 x 400 mL 96% D_2O minimal medium supplemented with 10% Spectra9 to O.D. 0.6, induce and harvested after 6 hours.

4.5. Production of a $^{2}H/^{13}C/^{15}N$ Sample for NMR Experiments

This section discusses the method and describes the purification of a $^{2}H/^{13}C/^{15}N$ labelled sample to be used in future NMR experiments to assist in the assignment of the **bb'x** backbone. The previous sections have discussed the development of methodology to maximise yield and labelling efficiency. In section 4.4.2 a final method was established which enabled the production of a sample with approximately 76% overall labelling and a concentration of 0.5mM from a 40 mL expression. This method of simple induction using minimal medium supplemented with 10 % Spectra9 rich medium was scaled up to 800 mL to produce sufficient quantities of protein for NMR experiments.

The 96% adapted cells were firstly streaked on a LB agar plate with appropriate antibiotics added; a colony was added to a small 50 mL 96% D_2O minimal medium supplemented with 10% Spectra9 and grown overnight. This small

starter culture was used to inoculate a larger 2x 400 mL 96% D_2O minimal medium supplemented with 10% Spectra9 at O.D. 0.1 and was grown to O.D. 0.6. The culture was then induced and cells harvested after 6 hours of induction. The cells were lysed using the freeze-thaw method and purified using affinity chromatography; samples were taken throughout the expression and purification and run on SDS-PAGE to follow the purification.

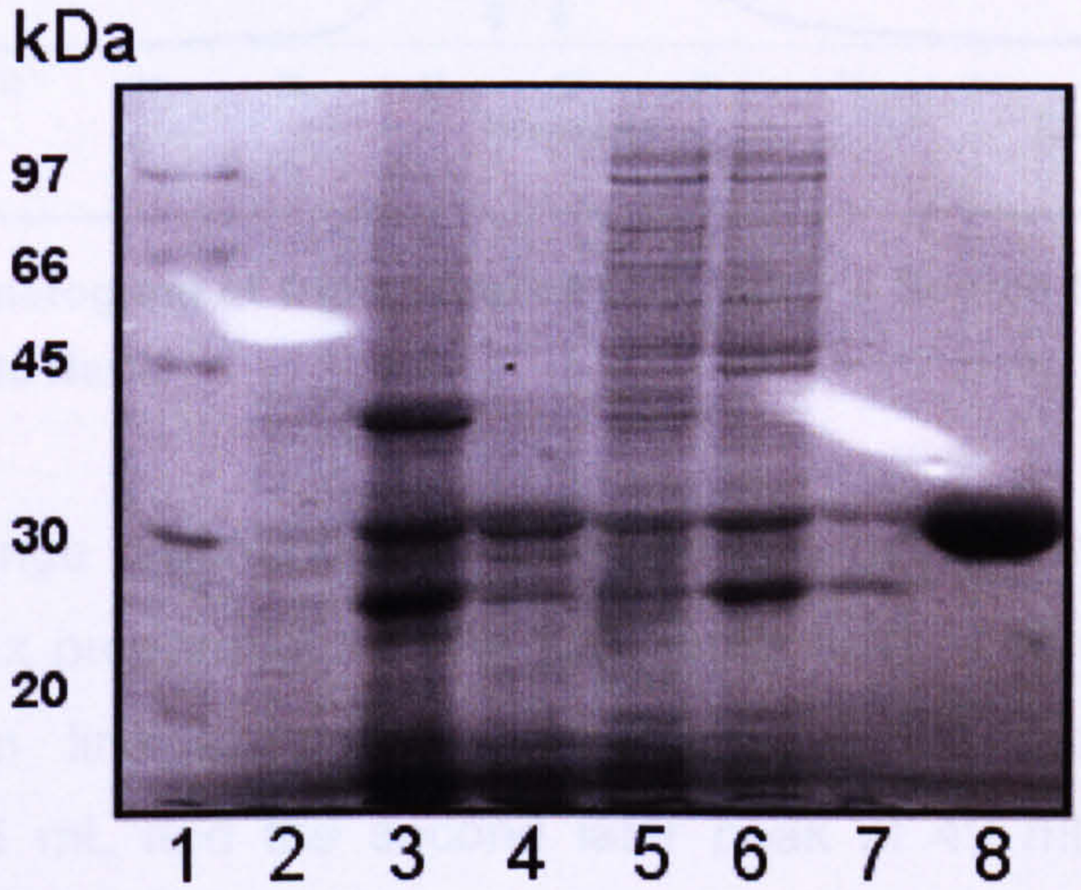

Figure 4-20. 16% SDS-PAGE 96% D_2O supplemented with 10% Spectra9 400 mL 6 hour expression. Lane 1, low molecular weight marker. Lane 2, soluble protein before induction. Lane 3, total protein after 6 hour induction. Lane 4, soluble protein after 6 hour induction. Lane 5, flow through. Lane 6, imidazole wash. Lane 7, low salt wash. Lane 8, hexa-histidine tagged triple labelled bb'x protein elute.

From the purification it was clear that a high level of expression was achieved, Figure 4-20 shows the purification of 400 mL (half the total expression was processed to maximise purification and reduce losses by overloading affinity column). Hence this purification was repeated with the other 400 mL expression culture (gel not shown; results were comparable). From the SDS-PAGE it was clear that purification was successful at removing the vast majority of protein impurities with little losses during the wash stages.

The two elutes were pooled and then diluted 5 fold in buffer A and applied to anion-exchange column for further purification using a salt gradient of 0 to 1M NaCl to elute the protein.

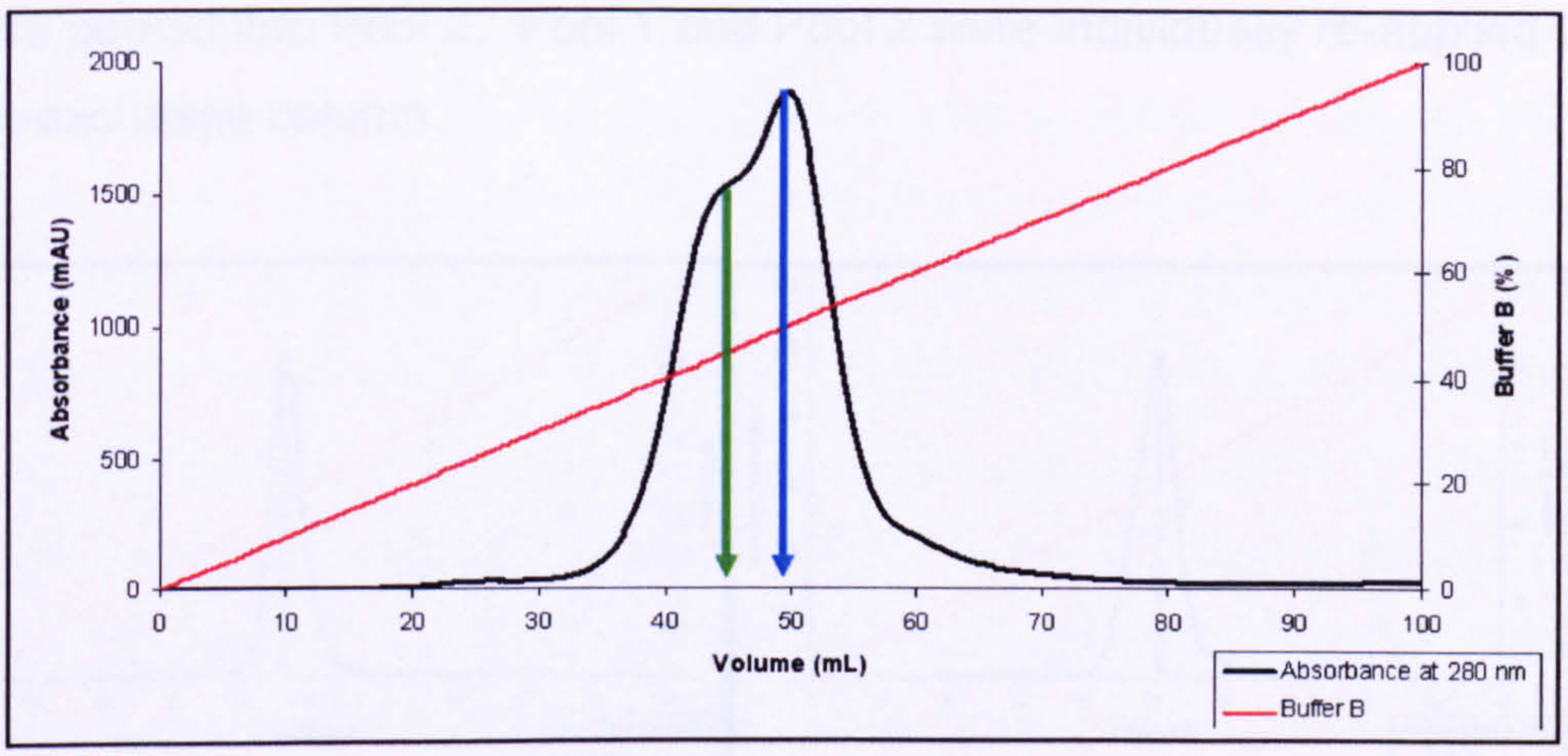

Figure 4-21. Chromatogram of triple labelled bb'x using a Source 30Q column with a 0-100 % gradient of 1M NaCl.

The anion-exchange chromatogram revealed two overlapped peaks, as seen with all other **bb'x** preparations. The two peaks have different maxima; in the figure the green line shows the maximum of the first peak which is approximately 45 mL and the second later peak at 49 mL. To identify the properties of the eluted protein peak, samples from across the peak were run on native-PAGE.

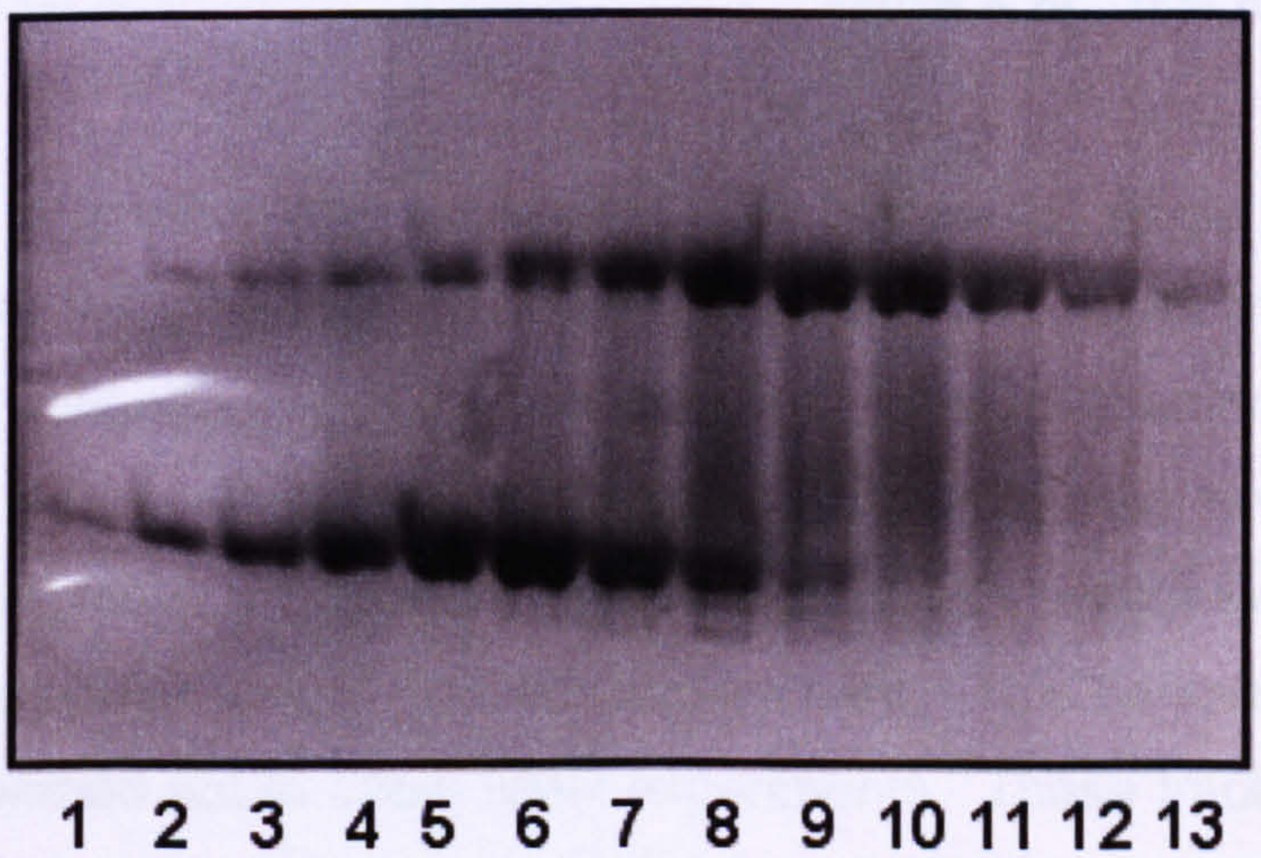

Figure 4-22. 16% Native-PAGE of samples taken across the peak in Figure 4-21.

The native-PAGE reveals the peak is made up of two species as seen in previous preparations of **bb'x**. Lanes 1 to 6 were pooled into Pool 1 and lanes

7 to 13 pooled into Pool 2. Pool 1 and Pool 2 were individually re-applied to the anion-exchange column.

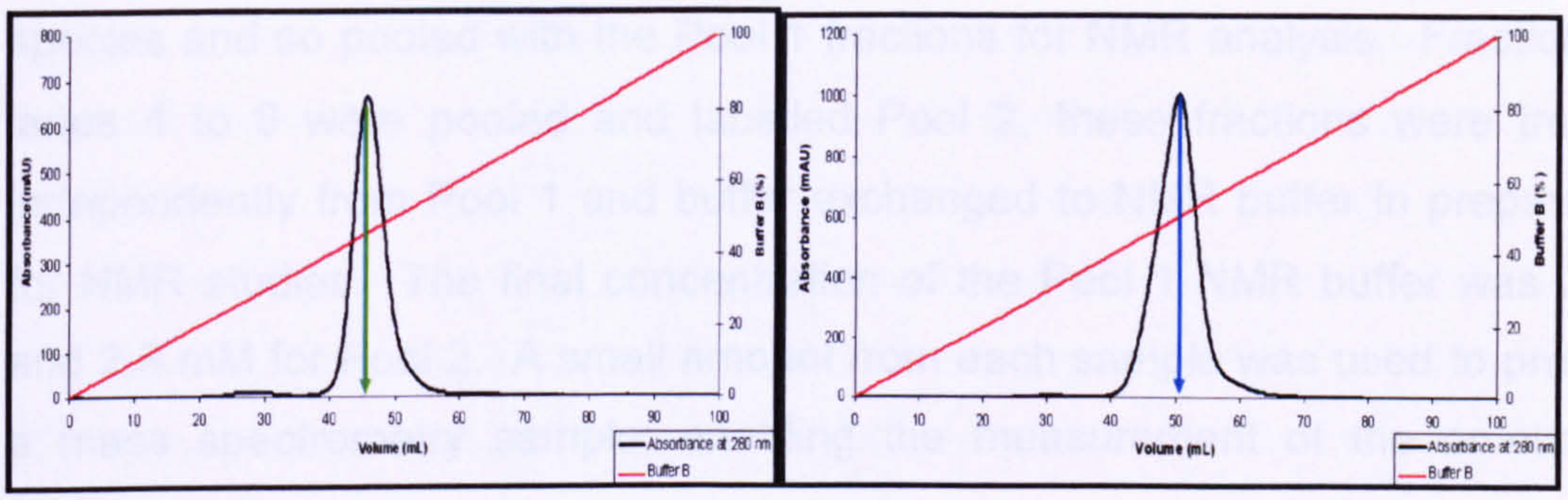

Figure 4-23. Anion-exchange chromatograph of Pool 2

Figure 4-24. Anion-exchange chromatograph of Pool 1

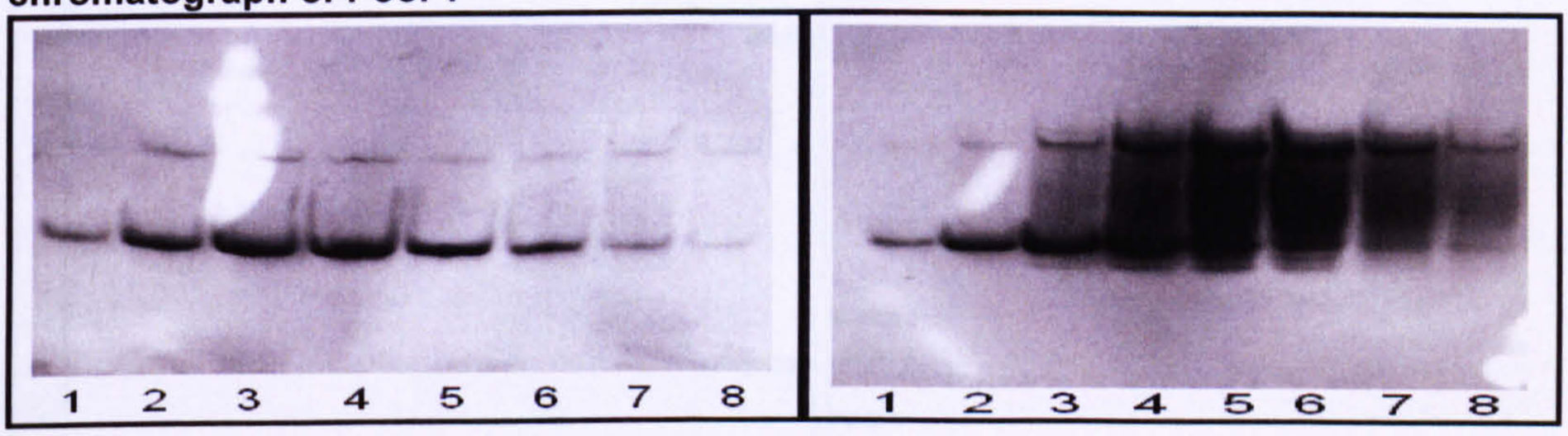

Figure 4-25. 16% native gel of Pool 1

Figure 4-26. 16% native gel of Pool 2.

The Pool 1 chromatogram in Figure 4-24 shows a single peak and as highlighted with the green line the maximum of the peak is 45 mL which elutes at an identical volume to the first peak identified in the initial anion-exchange run in Figure 4-21. The native-PAGE of fractions across the peak reveals the major species was the lower molecular weight in Figure 4-25, with a very low amount of the high molecular weight species contaminant. This contamination is very low and so deemed not to affect NMR experiments. These fractions were then pooled and buffer exchanged into the NMR buffer ready for further experiments.

The Pool 2 chromatogram in Figure 4-23 was also in agreement with the initial anion-exchange chromatogram in Figure 4-21, in that the maxima of the peaks

were both eluted around 50 mL; highlighted with the blue line. The native gel of the fractions of the peak in Figure 4-26, revealed a mixture of the two species. From the native-PAGE, lanes 1 to 3 were mainly the lower molecular weight species and so pooled with the Pool 1 fractions for NMR analysis. Fractions in lanes 4 to 8 were pooled and labelled Pool 2, these fractions were treated independently from Pool 1 and buffer exchanged to NMR buffer in preparation for NMR studies. The final concentration of the Pool 1 NMR buffer was 2mM and 2.8 mM for Pool 2. A small amount from each sample was used to prepare a mass spectrometry sample; enabling the measurement of the amount of isotope incorporation.

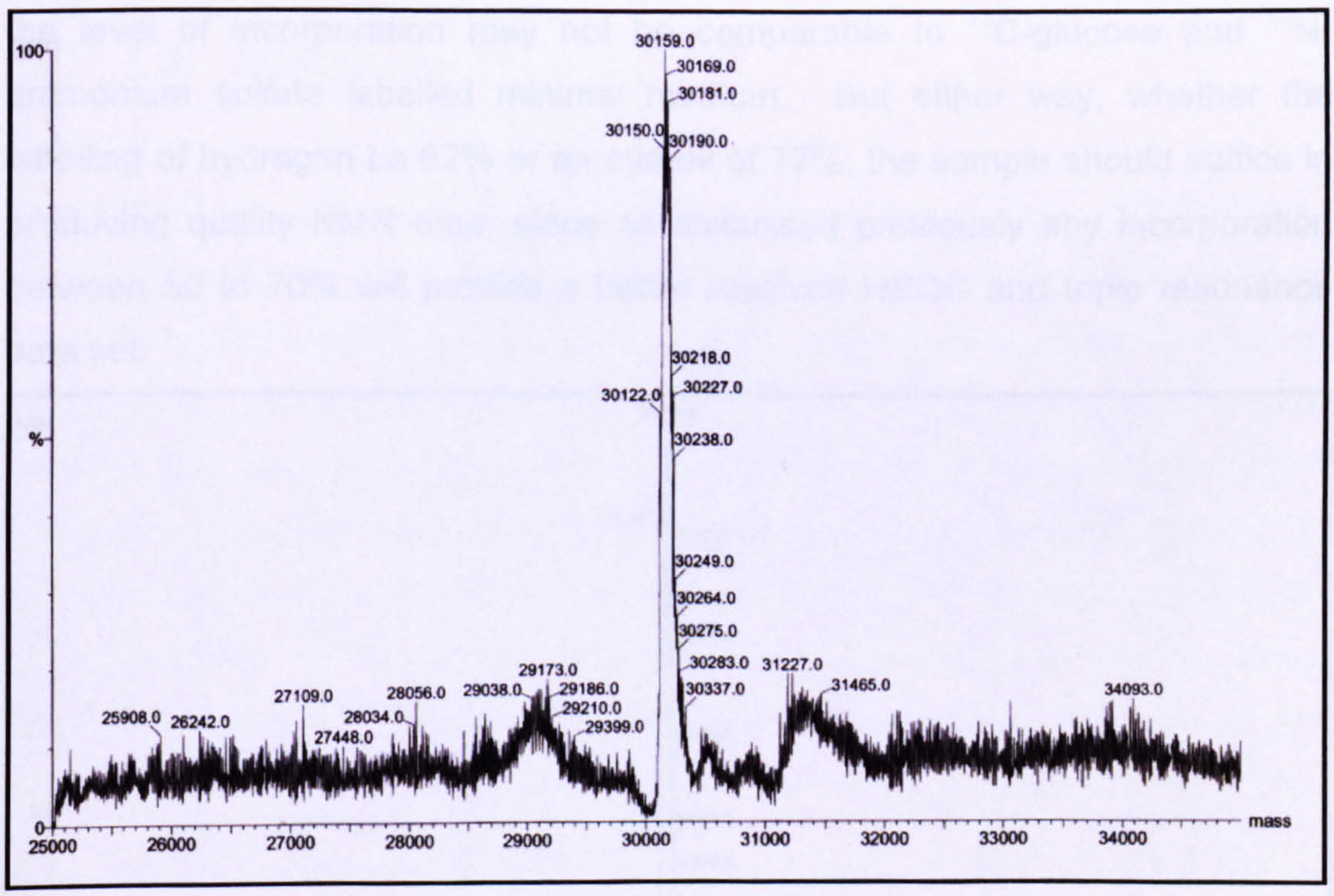

Figure 4-27. Positive ion mass spectrum of triple-labelled bb'x Pool 1.

There are many peaks highlighted by the mass spectrum deconvolution software, with a range of masses which are produced by species with different levels of deuterium incorporation. The mass spectrum of Pool 1 does reveal the major species of 30159 Da, hence from the unlabelled species of 27477 Da, 2682 atoms were labelled and so an overall labelling efficiency of 77.3% (3468

maximum number atoms of Carbon, Nitrogen and Hydrogen available to label). To be more accurate the double labelled **bb'x** sample was labelled with a 95.8% labelling efficiency described in section 3.3.4. So assuming the same level of efficiency was achieved of 95.8% of 1242 Carbon atoms and 316 Nitrogen atoms; then 1493 atoms would be labelled. Therefore, if 2682 atoms in total were labelled in Pool 1 then 2682-1493 atoms, would give the number of labelled hydrogen atoms of 1189, out of a possible 1910, hence a 62% labelling efficiency. This was based on the assumption the same level of Carbon and Nitrogen labelling would occur. This can not strictly be assumed since the medium was supplemented with Spectra9 $^{2}H/^{13}C/^{15}N$ labelled rich medium and the level of incorporation may not be comparable to ^{13}C-glucose and ^{15}N-ammonium sulfate labelled minimal medium. But either way, whether the labelling of hydrogen be 62% or an overall of 77%, the sample should suffice in producing quality NMR data, since as discussed previously any incorporation between 50 to 70% will provide a better resolved HSQC and triple resonance data set.

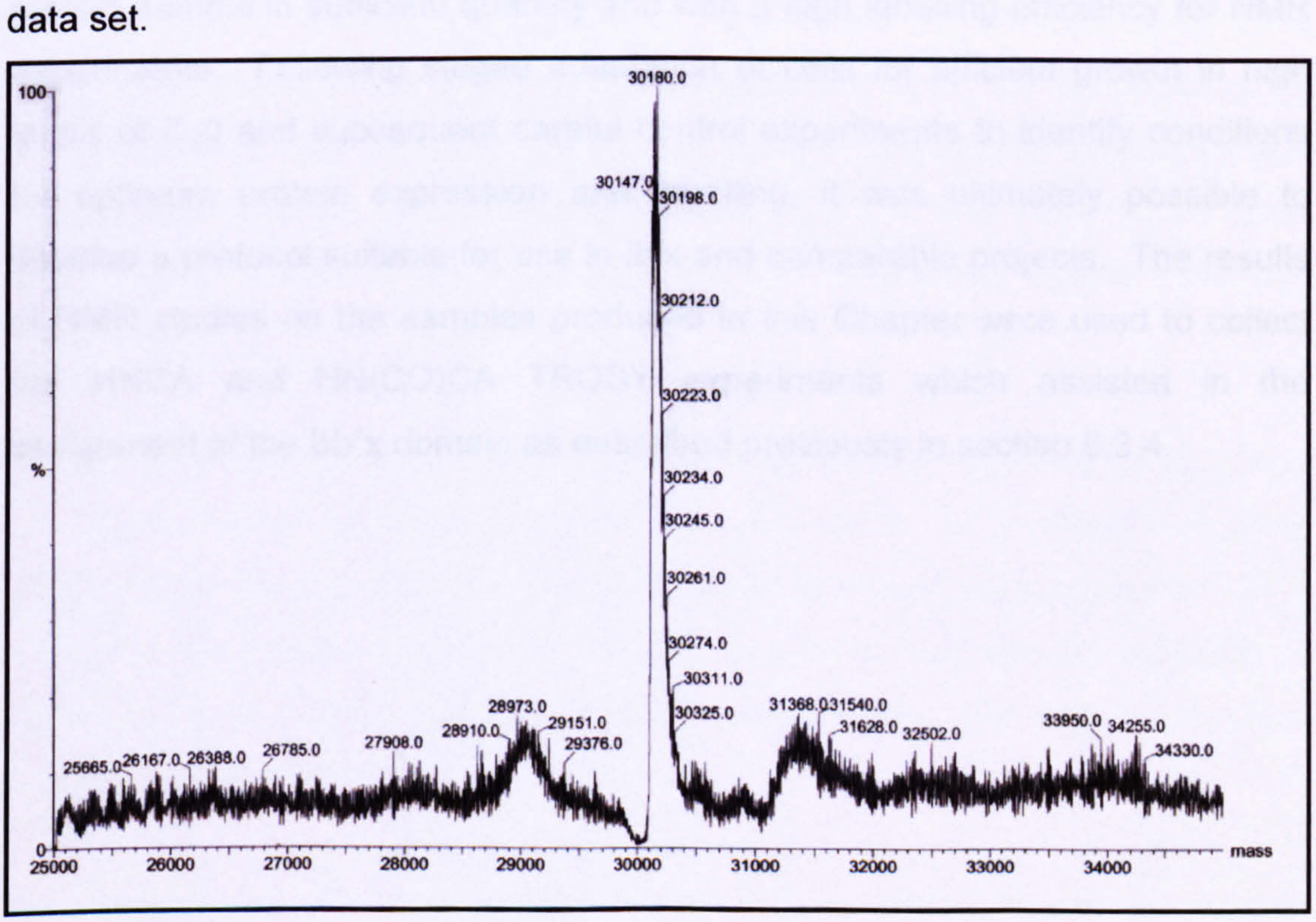

Figure 4-28. Positive ion mass spectrum of triple-labelled bb'x Pool 2.

The Pool 2 sample also reveals many species with varying molecular weights. This mass spectrum has a far greater resolution to any other spectra; as a result many peaks have been highlighted by the analysis software. The range of masses detected, as with the Pool 1 sample, was not deemed a problem, since these are relatively minor species. The main species has a molecular weight of 30180 Da, hence 2703 atoms have been labelled with an overall efficiency of 77.9%. Using the same assumption as Pool 1, whereby previously **bb'x** was double labelled with an efficiency of 95.8%, the number of hydrogen's labelled calculated using this method would be 1210, hence a 63.4% labelling efficiency.

4.6. Discussion

When the work described in this section was initiated, there were several uncertainties as to the optimum strategy for producing a $^{2}H/^{15}N/^{13}C$-labelled protein sample in sufficient quantity and with a high labelling-efficiency for NMR experiments. Following staged adaptation of cells for efficient growth in high levels of D_20 and subsequent careful control experiments to identify conditions for optimum protein expression and labelling, it was ultimately possible to develop a protocol suitable for use in this and comparable projects. The results of NMR studies on the samples produced in this Chapter were used to collect the HNCA and HN(CO)CA TROSY experiments which assisted in the assignment of the **bb'x** domain as described previously in section 6.3.4.

Chapter 5. Initial NMR Studies on PDI Domain Combinations

5.1. Introduction

This chapter has two sections, one which describes initial NMR studies carried out prior to this research project by Dr. R. Williamson (University of Kent) and the second describing new preliminary NMR studies carried out as part of this research project. These studies are used to explain the rationale behind attempting to assign the backbone structure of the **bb'x** domain construct as opposed to focussing on the **b'** domain only, when the **b'** domain is the main focus of this research project. The HSQC resolution of several domain combinations were compared and were used to validate a suitable methodology to obtain the **b'** backbone assignments. 2D-HSQC experiments serve as a quick experimental tool, run over approximately 1 hour, as an indicator of sample quality. Well resolved and concise resonances indicate that the protein sample was pure, the protein was stable and therefore was likely to give good quality data in further 3D NMR experiments required to obtain backbone assignments.

This chapter also describes how understanding the heterogeneity of **bb'x** and **b'x** preparations enabled progress to be made in structural studies. Fractionation of preparations to generate more homogenous samples for NMR improved the spectra significantly. It was also found that spectra collected at higher temperatures gave further improved spectra and could be used to allow full backbone assignments.

To increase the accuracy/confidence of **bb'x** assignment, the **b** domain chemical shifts were required and used to compare the **bb'x** chemical shifts;

hence overlapped peaks were assigned with a greater degree of accuracy/confidence. As a consequence of the newly identified advantage of sample fractionation, it was possible also to assign the backbone of the **b'x** domain; this data, as for the **b** domain, was used to increase the confidence in correct assignment of the **bb'x** domain.

5.2. Prior NMR Studies on PDI Domain Combinations

This section describes initial NMR studies carried out prior to this research by Dr. R. Williamson (University of Kent). General analysis in this section describing the quality of the spectra was done by comparing the peak intensity, resolution and broadening. Initial NMR studies carried out by Dr. R. Williamson; involved an analysis method of measuring the peak widths (in Hz) of the same 20 peaks (where possible) in HSQC spectra of different domain constructs.

Construct	Peak Width (Hz)
b'	>40
b' + KFWWFS	18.5
bb'	36.9
bb'x	24.0
b'x	26.1

Table 6. Summary of peak width analysis of different domain combinations. Same peaks used for all constructs, except bb' where a selection of peaks from b and b' domains were used. The b' is a conservative estimation due to difficult in measurements using the poor, peak-broadened spectrum. KFWWFS is a short hexa-peptide ligand.

5.2.1. b' domain

The **b'** domain was the main domain of focus in this research. Unfortunately the individual domain gave poor resolution in 2D-HSQC experiments. As shown in Figure 5-1 the spectrum was very poorly resolved since the peaks are

broadened and many are overlapped. The HSQC gives correlations between amide protons (^{1}HN) and amide nitrogen (^{15}N) frequencies in a single amide, so each residue except Proline residues will give one peak in the HSQC spectrum. The spectrum shows 87 peaks but the **b'** domain has 108 residues (not including the histidine tag or prolines), therefore 21 peaks are missing. The poorly resolved HSQC made the analysis of peak-widths difficult and so a conservative estimate of >40 Hz was given to the **b'** domain as reported in Table 6.

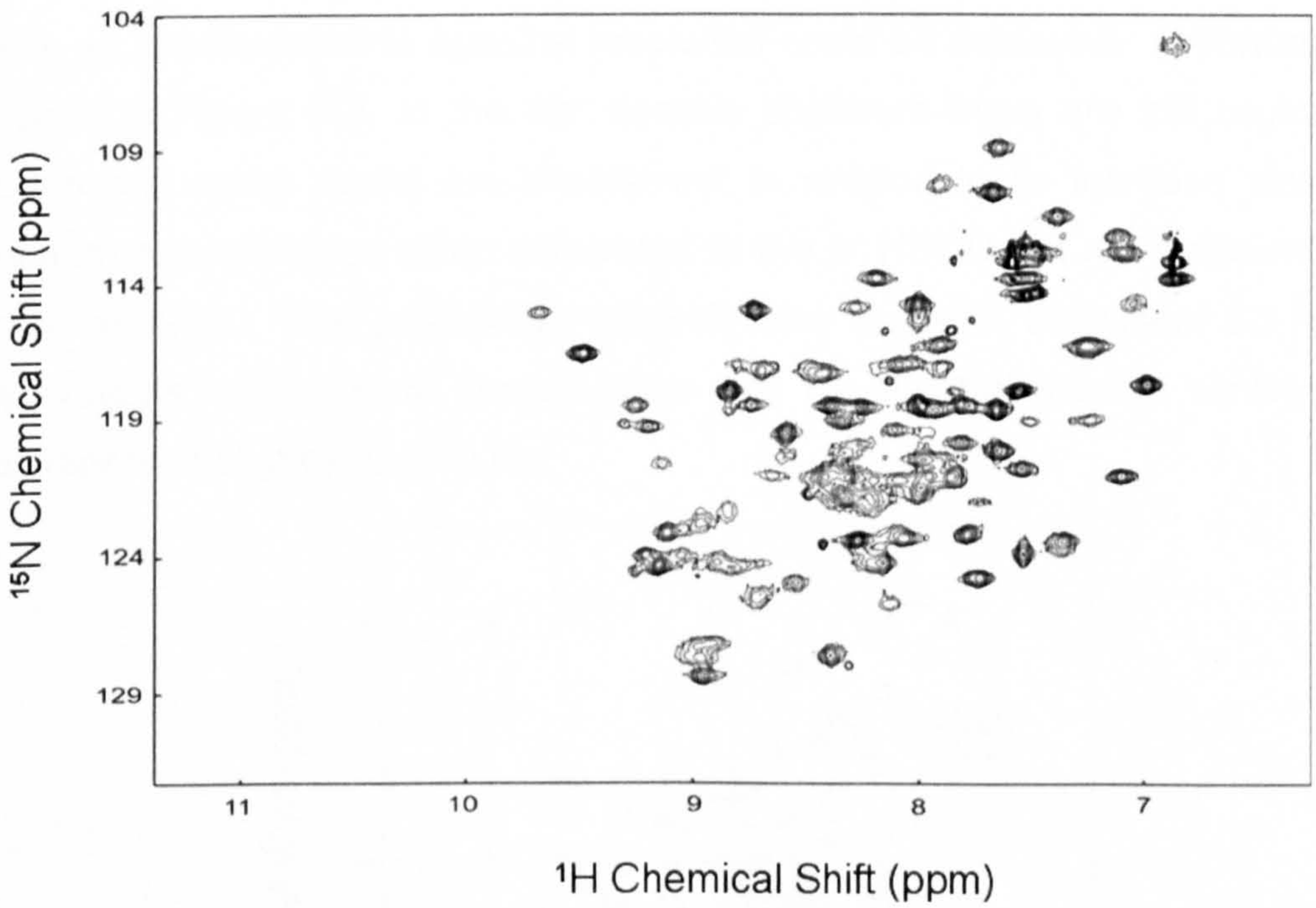

Figure 5-1. ^{15}N/^{1}H HSQC spectrum of ^{15}N-b' domain (sample and spectrum produced by Dr. R. Williamson).

The poor resolution of 2D-HSQC data could result from the **b'** domain being in conformational exchange; speculated to be an essential functional feature based on **b'** domain involvement with substrate binding. In one experiment, the spectrum was dramatically improved by the addition of a short hexa-peptide ligand (KFWWFS) as highlighted in Table 6, where the average peak-widths of

>40Hz decreased to 18.5 Hz; but unfortunately this result has not been possible to reproduce. Due to the low resolution of the **b'** domain, alternative domain combinations were studied that could lead to better NMR resolution, by stabilising the **b'** domain.

5.2.2. bb' domain

The **bb'** domain construct was analysed by 2D-HSQC experiments, it was hoped that with the addition of the **b** domain, the **b'** domain would be stabilised; hence an improvement in spectral resolution could be achieved. Unfortunately, as seen in Figure 5-2, in the **bb'** domain spectrum there are still regions of overlap and many peaks are broadened; in relation to its increase size, the spectrum has improved when compared to the **b'** HSQC but not sufficiently for further analysis. The peak-width analysis also supports the visual 2D-HSQC observations, since the **b'** domain gave >40 Hz, the **bb'** domain combination was improved slightly to 36.9 Hz.

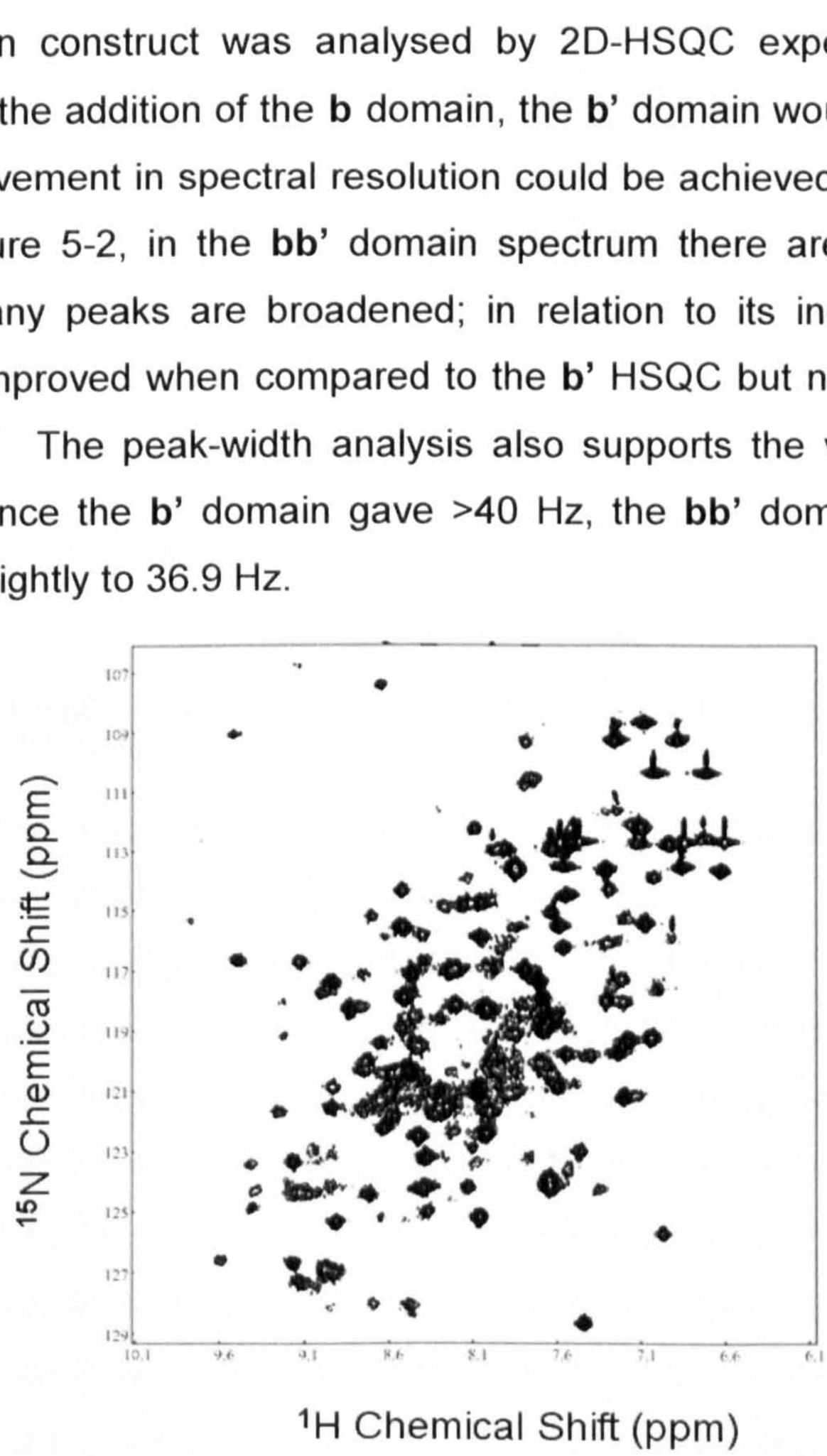

Figure 5-2. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb' domain (sample and spectrum produced by Dr. R. Williamson).

5.2.3. bb'x domain

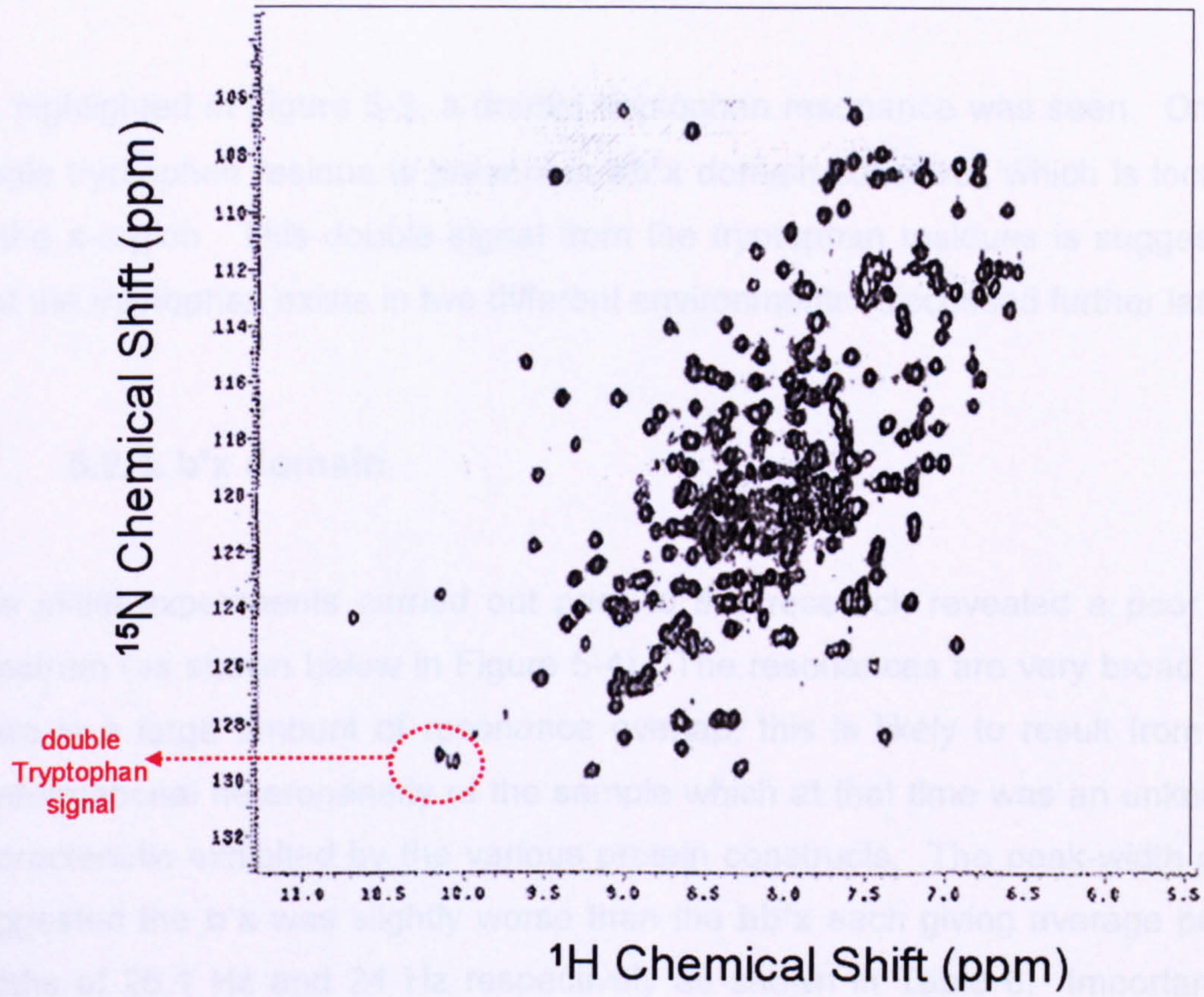

Figure 5-3. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb'x domain Pool 1 (sample and spectrum produced by Dr. R. Williamson).

The **bb'x** samples in these early experiments were crudely fractionated on the basis of the two peaks identified during anion-exchange chromatography purification and not using native-PAGE analysis. The 2D-HSQC run on the pooled fraction from the early anion exchange peak (Pool 1) as shown in Figure 5-3, shows that the resolution was again an improvement on the **b'** domain and **bb'** domain construct. The vast majority of the peaks are well resolved and concise; there is some peak broadening and some degree of overlap resulting from the high number of residue signals given by this larger domain construct. The peak-width analysis details the extent of the improvement, whereby this **bb'x** Pool 1 sample gave peak-widths of 24 Hz whereas the peak-widths of the **b'** and **bb'** domains were much higher (see Table 6). Pool 2 gave very poor spectra (data not shown) again consistent with the theory that the late pool

(Pool 2) is in a more flexible conformation; indicated by broad/poorly resolved resonances.

As highlighted in Figure 5-3, a double tryptophan resonance was seen. Only a single tryptophan residue is present in **bb'x** domain construct, which is located in the **x**-region. This double signal from the tryptophan residues is suggestive that the tryptophan exists in two different environments (discussed further later).

5.2.4. b'x domain

The initial experiments carried out prior to this research revealed a poor **b'x** spectrum (as shown below in Figure 5-4). The resonances are very broad and there is a large amount of resonance overlap, this is likely to result from the conformational heterogeneity of the sample which at that time was an unknown characteristic exhibited by the various protein constructs. The peak-width data suggested the **b'x** was slightly worse than the **bb'x** each giving average peak-widths of 26.1 Hz and 24 Hz respectively as shown in Table 6. Important to note, is that the sample of **b'x** used here was not stringently fractionated, so an improvement was later found to be possible with fractionated samples as shown to be possible with **bb'x** samples in section 5.3.1. The increased resolution achievable by fractionation is highlighted by comparing the spectrum collected for an unfractionated sample of **b'x** shown in Figure 5-4 and a fractionated **b'x** shown in the later Figure 5-7.

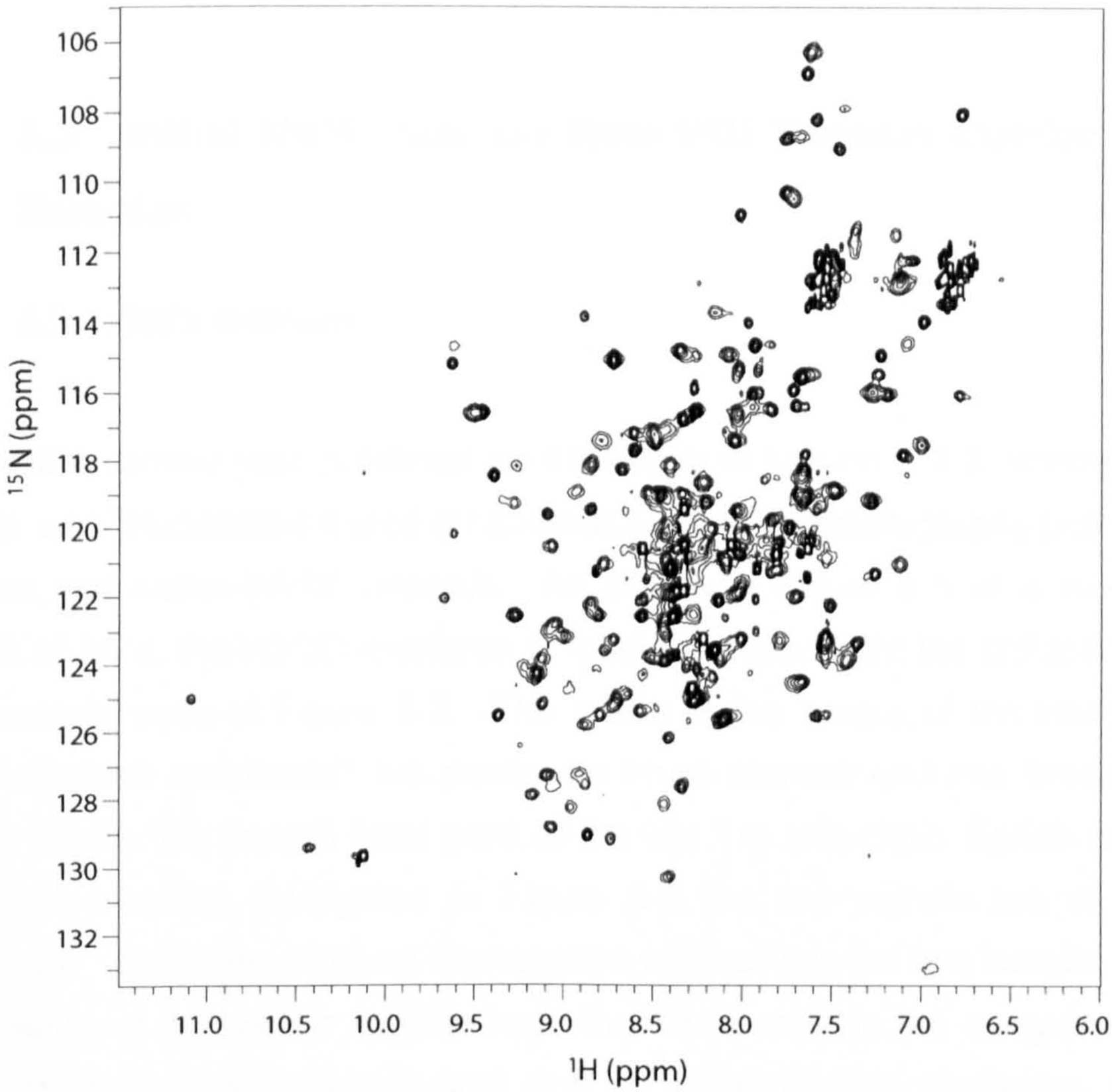

Figure 5-4. $^{15}N/^{1}H$ HSQC spectrum of unfractionated ^{15}N-b'x domain (sample and spectrum produced by Dr. R. Williamson).

From Table 6 it is clear there is improvement in the spectral quality with the addition of the **b** domain and **x**-region to domain combinations. This is most clearly visible by the decrease in the average peak-widths of **bb'** (36.9 Hz) and **bb'x** (24.0 Hz). The method of how the **x**-region was able to stabilise the whole domain construct, particularly the **b'** domain, is not known but is speculated to be as a result of the **x**-region binding to the **b'** domain active site and so increasing the **b'** domain conformational stability. At this stage the most promising domain combination with the potential to enable full backbone assignment was the **bb'x** as it gave the lowest spectral peak-width measurements.

5.3. Initial NMR Data on New PDI Domain Combination Samples

5.3.1. bb'x domain

A ^{15}N-**bb'x** sample was produced as described in section 3.3.3, whereby the sample was fractionated based on anion-exchange chromatography peaks, gel filtration and native-PAGE analysis. As shown in Figure 5-5 of a monomer sample of **bb'x**, the HSQC resolution is much improved from the unfractionated **bb'x** sample used in Figure 5-3. The peaks in the centre of the HSQC are better resolved and overall, the peaks are much sharper and less broadened. Also in Figure 5-5, the left hand peak of the two Trp side-chain signals is more prominent whereas highlighted in Figure 5-3 the two signals are of equal magnitude. The more stringent fractionation process has led to a sample where the majority of the protein species have the Trp residue in the **x**-region buried (likely to be in a hydrophobic site) and so stabilises this conformer, hence producing a well resolved spectrum, as seen with **b'x** mentioned previously.

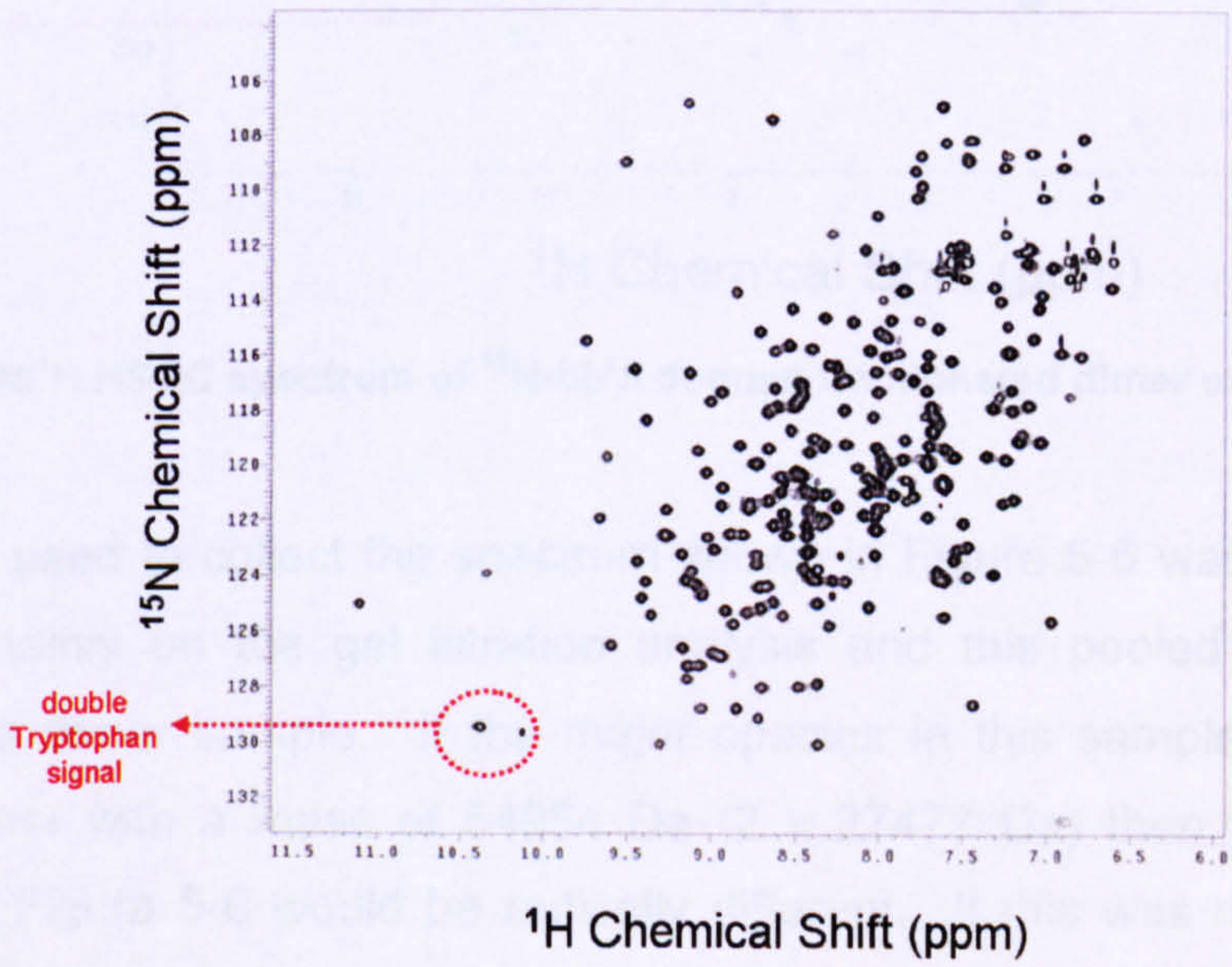

Figure 5-5. ^{15}N/^{1}H HSQC spectrum of ^{15}N-bb'x domain fractionated monomer sample.

The fractionated dimer sample gave a very poor HQSC spectrum as shown in Figure 5-6, all the resonances are broadened and there is a large amount of overlapping. This supports the theory of two conformations, whereby the dimer sample is a less structurally stable conformer. Also as highlighted in Figure 5-6, the two signals from the Trp side chain have almost disappeared; indicating a lack of **x**-region structural stability. The **x**-region is likely to be unbound hence the Trp residue is not buried in a hydrophobic site as seen with monomer sample in Figure 5-5.

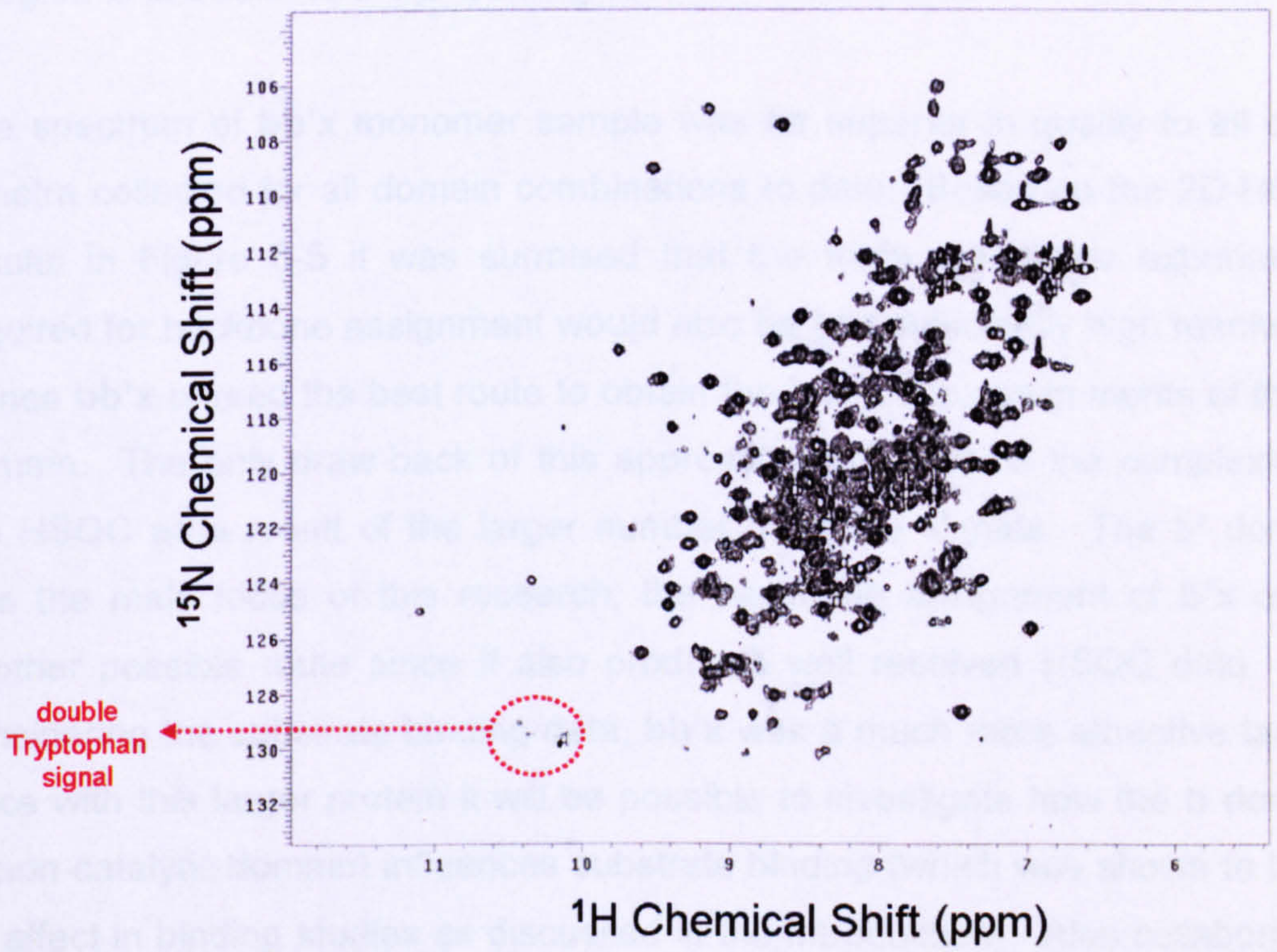

Figure 5-6. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-bb'x domain fractionated dimer sample.

The sample used to collect the spectrum shown in Figure 5-6 was fractionated based on mainly on the gel filtration analysis and this pooled sample was defined as a dimer sample. If the major species in this sample was a non-covalent dimer with a mass of 54954 Da (2 x 27477 Da) then the spectrum observed in Figure 5-6 would be radically different. If this was the case then very little would be observed, a 55kDa protein sample would have massive peak

broadening and with a size of 55kDa very few peaks if any would be observed and the spectrum would appear smeared rather than having defined peaks; hence the spectrum in Figure 5-6 is not representative of a sample containing primarily a dimer species. The spectrum indicates a less structured monomer species as opposed to a dimer species. It was therefore hypothesised based also on the observations of the **x**-region tryptophan residue, HSQC spectrum and AUC analysis (in Section 3.8) that the dimer fraction was a mixture of a small amount of non-covalent dimer species and a monomer species where the **x**-region is unbound to the **b'** binding site.

The spectrum of **bb'x** monomer sample was far superior in quality to all other spectra collected for all domain combinations to date. Based on the 2D-HSQC results in Figure 5-5 it was surmised that the triple resonance experiments required for backbone assignment would also be of a sufficiently high resolution. Hence **bb'x** offered the best route to obtain the backbone assignments of the **b'** domain. The only draw-back of this approach using **bb'x** is the complexity of the HSQC as a result of the larger number of amide signals. The **b'** domain was the main focus of this research; the backbone assignment of **b'x** offers another possible route since it also produces well resolved HSQC data. But considering the substrate binding data, **bb'x** was a much more attractive target, since with this larger protein it will be possible to investigate how the **b** domain (a non-catalytic domain) influences substrate binding (which was shown to have an affect in binding studies as discussed in the introduction). Also collaborative work with Dr. L. Ruddock on obtaining the x-ray crystal structure of the **b'x** domain was underway hence **bb'x** was focused on to provide additional information. Since the **b** domain had been solved by NMR an opportunity to reduce the difficulties in backbone assignment existed, despite the complexity of the **bb'x** HSQC spectrum. This involved firstly assigning the backbone structure of the **b** domain. Although the **b** domain has been solved, our running conditions are different and also there are always slight differences in spectral data using different NMR machines.

Since the **b'x** spectrum was almost as good in terms of quality, as that collected for **bb'x**, the **b'x** assignment could be attempted if **bb'x** assignments were found to be difficult and so assist with **bb'x** assignments in a similar fashion as the **b** domain.

5.3.2. b'x domain

The **b'x** sample now used was fractionated based on gel filtration and native-PAGE analysis and gave a much improved HSQC spectrum although some peaks are overlapped and broadened, as highlighted in blue in Figure 5-7. This spectrum is significantly better than **b'** and **bb'** spectra, most of the peaks are better resolved, indicating the **b'x** domain is more stable, with the addition of the **x**-region. Interestingly, the **b'x** domain contains one Trp residue, located in the **x**-region, but as highlighted in red in Figure 5-7, two signals from the Trp side chain are visible. This would suggest the **x**-region is in a slow exchange between two (or more) conformations (Nguyen *et al.*, (Manuscript in preparation)), see Appendix 1.

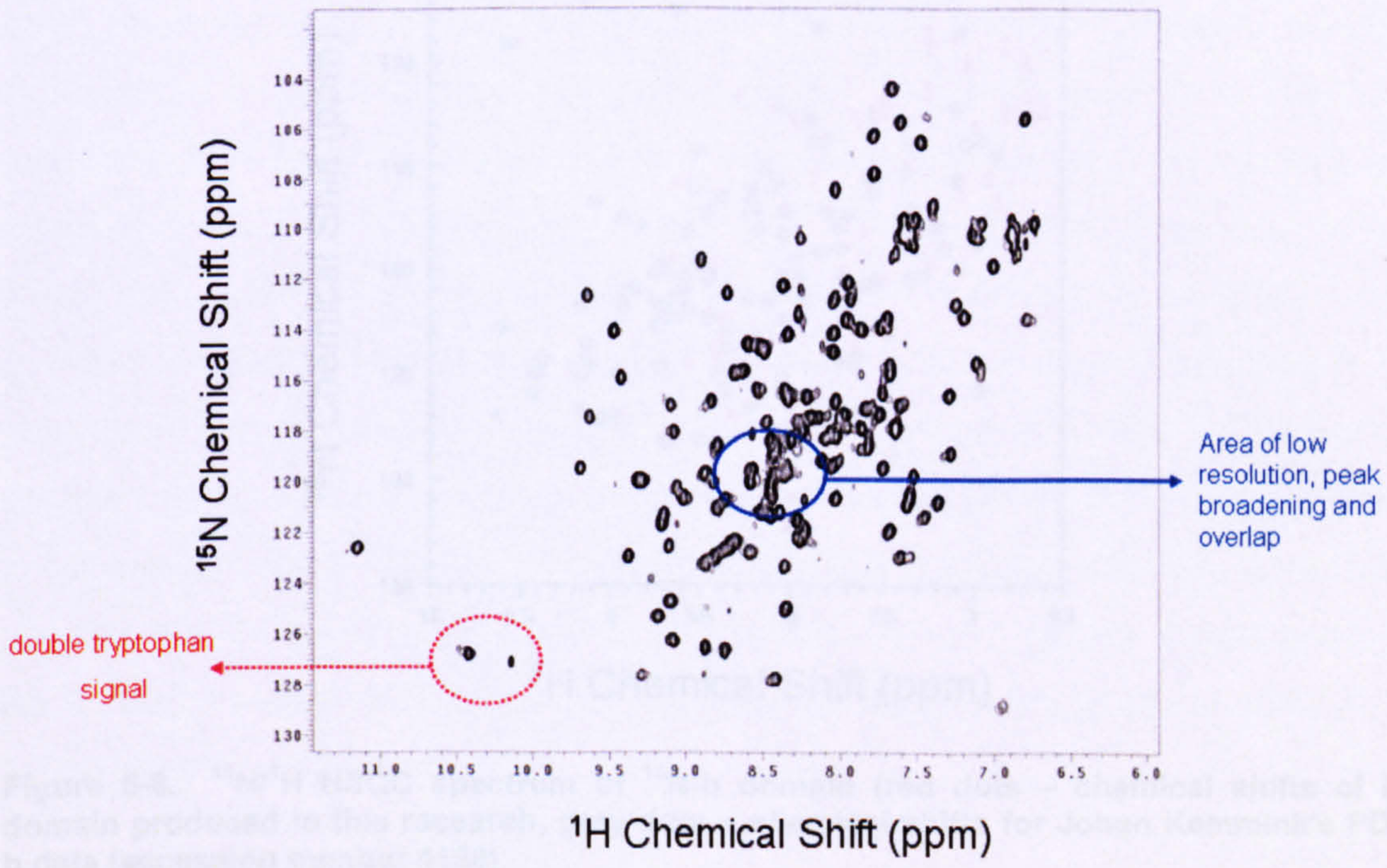

Figure 5-7. ^{15}N/^{1}H HSQC spectrum of ^{15}N-b'x domain monomer sample.

As seen here in Figure 5-7, the inclusion of the **x**-region resulted in the improvement of the HSQC, speculated to result from **x**-region binding to the **b'** domain binding site.

5.3.3. b domain

To aid the assignment of the **bb'x**, the **b** domain can be used to assist in accurate assignment. The already solved **b** domain chemical shifts can be used to quickly assign the **b** domain HSQC spectrum produced under our running conditions (slight differences in chemical shifts will exist as highlighted in Figure 5-8). The **b** domain was fully assigned using Johan Kemmink's PDI **b** data (Kemmink *et al.*, 1999) which were used to cross reference assignments. Furthermore to the 2D HSQC, TOCSY data was collected by Dr. R. Williamson then analysed and assigned by Dr. Katrine Wallis. With the full **b** domain backbone assignments, the assignment of **bb'x** was hoped to be simplified.

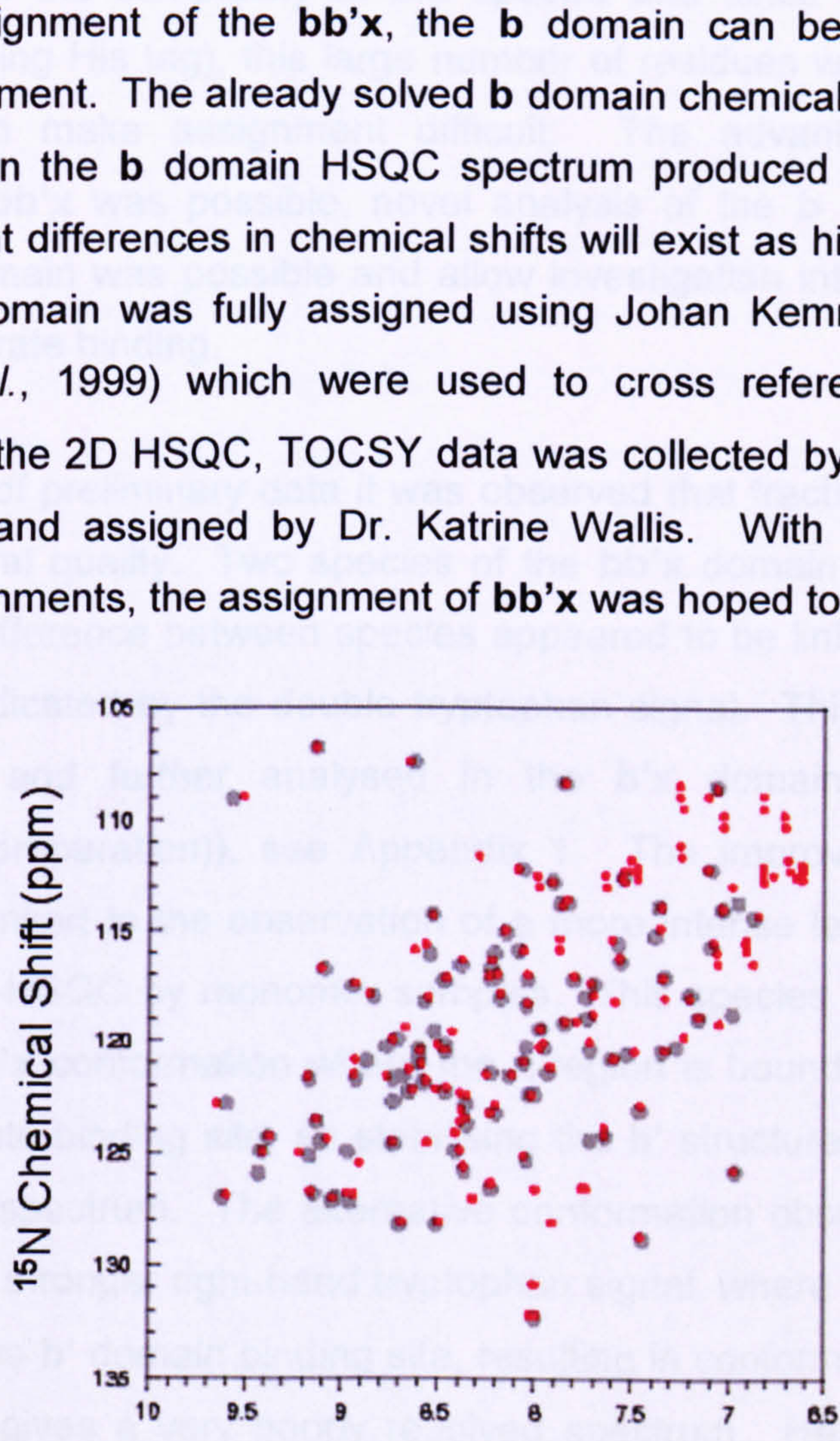

Figure 5-8. $^{15}N/^{1}H$ HSQC spectrum of ^{15}N-b domain (red dots – chemical shifts of b domain produced in this research, grey dots – chemical shifts for Johan Kemmink's PDI b data (accession number 4156).

5.4. Discussion

The **b'** domain was the main domain of interest, but spectral data was very poor. Spectral resolution was improved with the addition of the **b** domain and further improved with the addition of the **x**-region. This led to the conclusion that the **bb'x** domain offered the best route to obtaining **b'** assignments. The obvious drawbacks were the complexity of the spectra and since the **bb'x** was 249 residues (including His tag), this large number of residues would result in peak overlap and so make assignment difficult. The advantage would be, if assignment of **bb'x** was possible, novel analysis of the **b** domain interaction with the **b'x** domain was possible and allow investigation into the role of the **b** domain in substrate binding.

By comparison of preliminary data it was observed that fractionation also led to improved spectral quality. Two species of the **bb'x** domain combination were observed; the difference between species appeared to be linked to the **x**-region environment, indicated by the double tryptophan signal. This phenomena was also observed and further analysed in the **b'x** domain (Nguyen *et al.*, (Manuscript in preparation)), see Appendix 1. The improvement to spectral resolution was linked to the observation of a more intense left-hand tryptophan signal in the 2D-HSQC by monomer samples. This species was speculated to result from a **bb'x** conformation where the **x**-region is bound, speculated to be to the **b'** substrate binding site, so stabilising the **b'** structure and resulting in a much improved spectrum. The alternative conformation observed in the dimer samples gives a stronger right-hand tryptophan signal, where the **x**-region is not in contact with the **b'** domain binding site, resulting in conformationally un-stable species and so gives a very poorly resolved spectrum. Hence a fractionated **bb'x** sample was shown to offer a route with the highest probability of achieving full backbone assignment.

Chapter 6. NMR Backbone Assignment of bb'x

6.1. Introduction

This Chapter describes the approach to obtain the backbone resonance assignments of **bb'x**, using a double-labelled ($^{13}C/^{15}N$) **bb'x** sample. This sample was expressed and purified as described in section 3.3.4, whereby the sample was fractionated based on anion-exchange, gel-filtration and native-PAGE analysis; the improvement of resolution in 2D-HSQC experiments is described in section 5.3.1. Here this sample was used to generate triple resonance data, specifically, HNCACB and CBCA(CO)NH spectra. These data sets were used to obtain the sequence specific backbone resonance assignments (N, NH, C_α, C_β), as described in section 1.4.4, from the identification of the intra and inter-residue connectivities using the CcpNMR Analysis package (Vranken *et al.*, 2005).

Also as previously discussed, due to the large size of **bb'x**, the complexity of the data could make assignment difficult. Hence this chapter also describes how the **b** domain HSQC assignments were used to cross-reference the assignments on to the **bb'x** HSQC. It was also found to be necessary to assign the **b'x** backbone and also use these assignments to cross-reference on to the **bb'x** data using the same principle as with the **b** domain. Later it became necessary to produce a $^{2}H/^{13}C/^{15}N$ **bb'x** sample which was used to collect TROSY data and this provided another data set to aid the assignment. This three-way cross reference method allowed full unambiguous assignment of the **bb'x** and **b'x** with high level of confidence.

6.2. First Attempt at bb'x Backbone Assignment and Stability Analysis

Initial 2D-HSQC experiments were carried out at 25 °C. Although the sample gave good HSQC spectra, unfortunately this was not true in triple resonance experiments and so this section describes how the resolution was increased by raising the temperature at which the experiments were run.

6.2.1. Backbone assignment of bb'x at 25 °C

A 2D-HSQC experiment was run to test the quality of data using a double-labelled ($^{13}C/^{15}N$) **bb'x** sample, produced as described in section 3.3.4, before attempting triple resonance experiments. The assessment of quality was made on the basis of the two signals from the Trp indole side chain. As previously described a strong left hand peak, appeared stronger in Pool 1 or monomer fractions and was associated with high resolution spectra. As highlighted in Figure 6-1 of the $^{13}C/^{15}N$ **bb'x** sample and as seen previously with ^{15}N-labelled **bb'x** test spectra in Figure 5-5, the left hand indole resonance is the stronger peak. Also all the other resonance peaks are very clear and concise, there appears to be very little peak broadening. The overall quality is perhaps a little lower as there appears to be some overlap of resonances in the centre of the spectrum in comparison with the spectrum collected with the ^{15}N-labelled **bb'x** test spectra in Figure 5-5. This could result from a slight contamination with Pool 2, likely to have occurred during fractionation. But most of the peaks are well resolved and clear, therefore it was not deemed at this stage to be a serious problem.

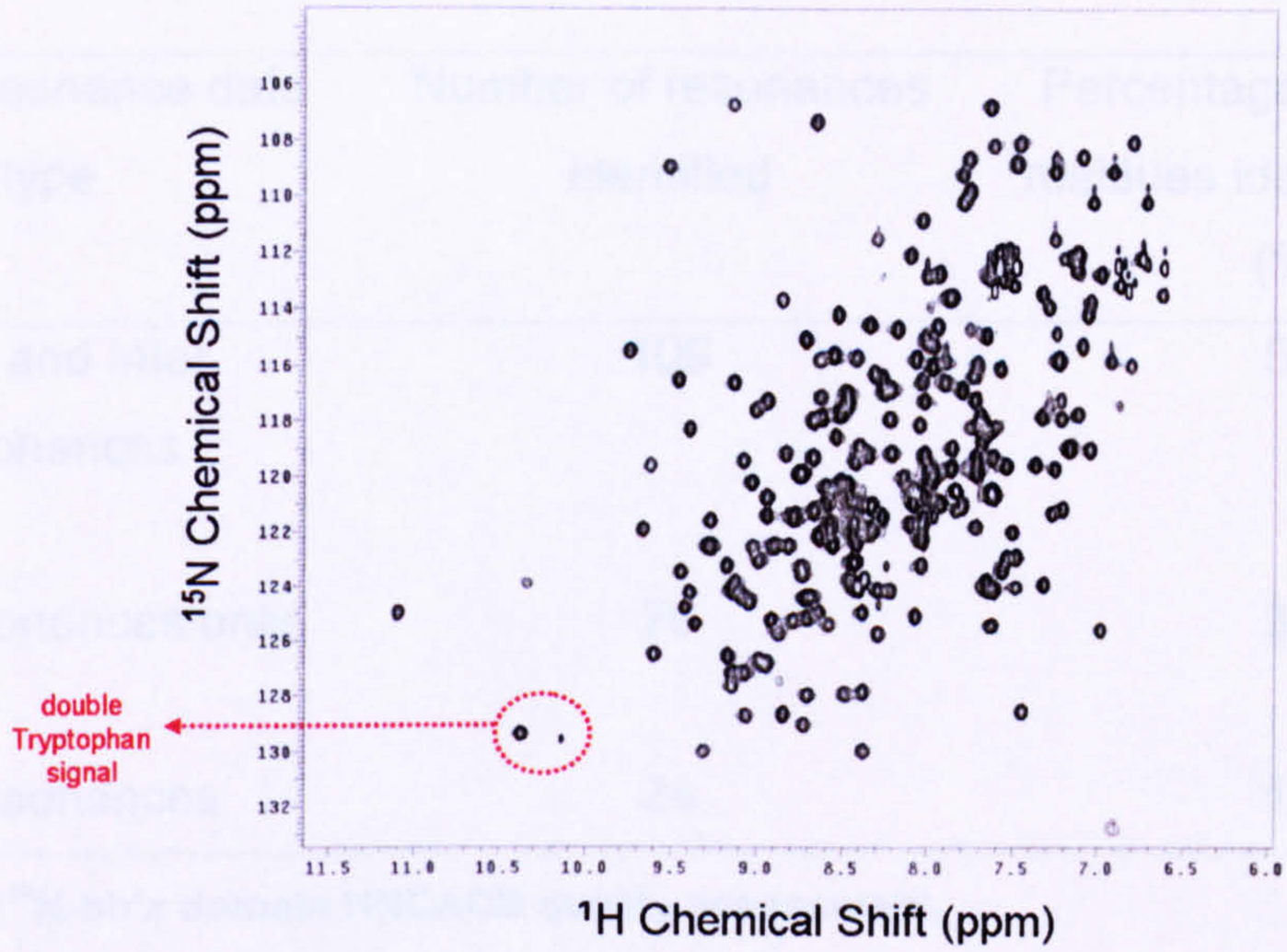

Figure 6-1. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 25°C.

The first triple experiment run was the HNCACB, which give the intra and inter-residue amide C_α and C_β resonances. A quick cursory analysis on this data set was carried out to investigate the quality of data obtained. 203 residues were identified in the 2D-HSQC and for each HSQC peak, the triple resonance was recorded in terms of whether all the data was visible (intra and inter resonances), half the data visible (Intra resonances only) or no data visible (No resonances). This data is presented in Table 7, from which it is clear that only 54% of quality usable data had been collected from this sample. This was far too low, even with the assistance of extra data sets; a full assignment would not be possible and a method of increasing triple resonance data quality was required.

Triple Resonance data type	Number of resonances identified	Percentage of the 208 residues identified seen (%)
Intra and Inter resonances	109	54
Intra resonances only	70	34
No resonances	24	12

Table 7. $^{13}C/^{15}N$-bb'x domain HNCACB quality assessment.

6.2.2. Stability experiment at 35°C

One method of improving NMR data quality and resolution would be to increase the temperature at which data was collected. The increase in temperature results in an increase in the molecular tumbling rate; this reduces the spectral linewidths and so increases spectral resolution. A 2D-HSQC was collected on the same sample that was used in section 6.2.1, to test the protein stability at 35°C. The HSQC was collected after 4 days at 35 °C, the approximate amount of time required to collect the HNCACB data set. The increase in temperature appeared not to affect the stability of the protein as shown in Figure 6-2, as the peaks were still concise and there was little peak broadened indicating the protein was still folded. However, some new peaks have begun to appear indicating some degradation of the protein sample and some of the peaks have weakened but remain visible. Unfortunately, the two signals from the Trp indole side chain now appear with equal intensity, perhaps indicating an equilibrium between the two conformers inferred here from the increase of the Pool 2 dimer (right hand indole) signal.

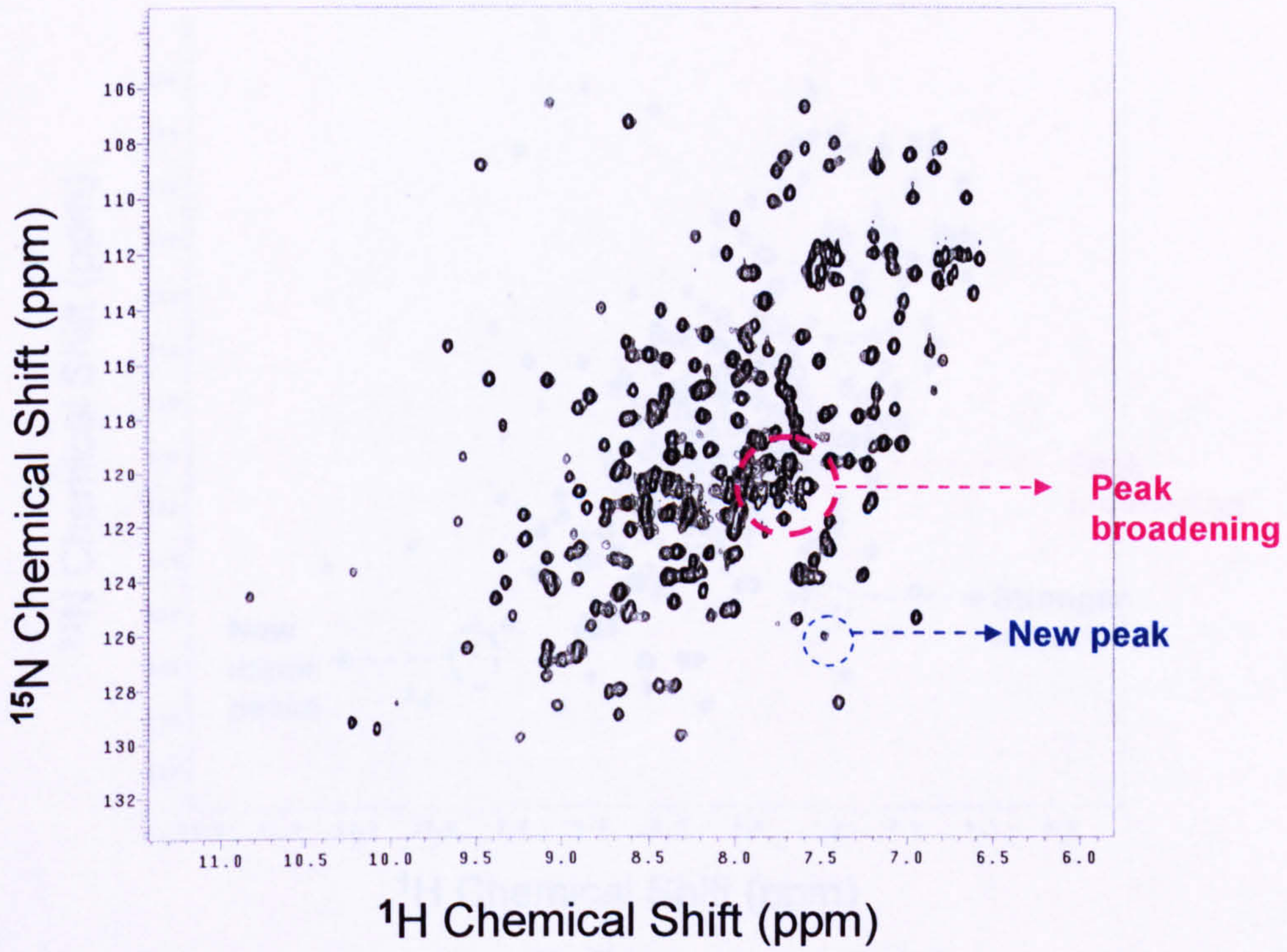

Figure 6-2. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 35°C after 4 days.

6.2.3. Stability experiment at 40°C

The NMR machine running temperature was then the raised to a maximum of 40°C, to determine whether the protein was stable at this temperature. Since the protein appears to be stable at 35°C and to maximise the improvement to resolution it was hoped data could be collected using the maximum temperature possible. A 2D-HSQC experiment was run 30 minutes after increasing the temperature to 40°C as shown in Figure 6-3. Generally the peaks are still concise indicating the majority of the protein sample was still folded, some more minor peaks have appeared indicating some further degradation; but overall the protein was still folded and stable.

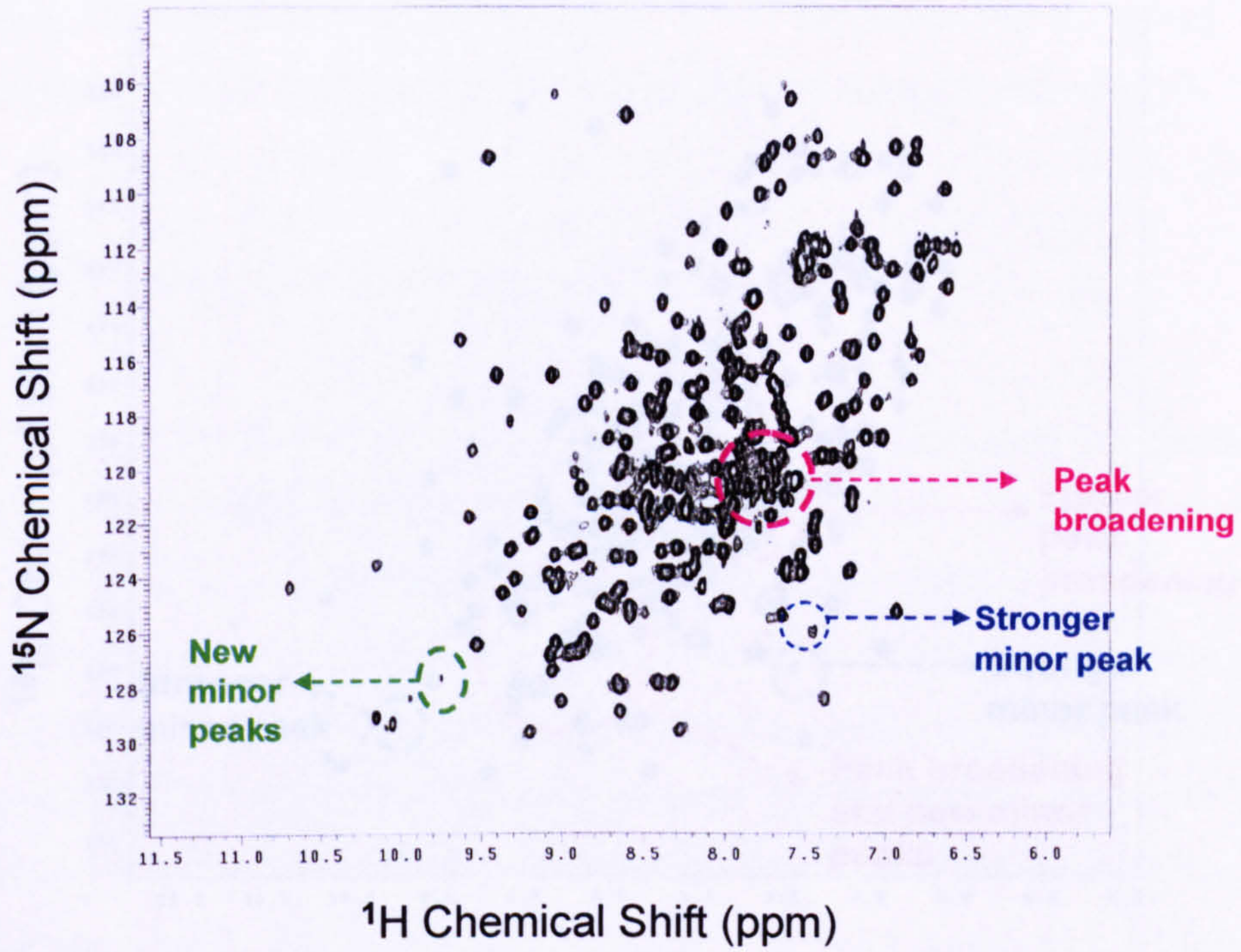

Figure 6-3. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 40°C after 30 minutes.

The protein sample remained at 40°C for 22 hours to further measure protein stability at 40°C as shown in Figure 6-4. Again the majority of the protein sample appears to remain in a folded state, since the peaks have not massively broadened. Some further degradation has taken place, as previously identified minor peaks seen in Figure 6-2 and Figure 6-3 become stronger and more new minor peaks appear. To summarise the treatment of this sample, it has been at 25°C for 4 days, 35°C for 4 days and 40°C for 1 day. This exceeds the normal run time for the collection of HNCACB and CBCA(CO)NH spectra which would be 8 days. So most importantly the protein sample appears to remain relatively well stable in these conditions and using a new double labelled protein sample it appears viable to collect triple resonance data at 40°C.

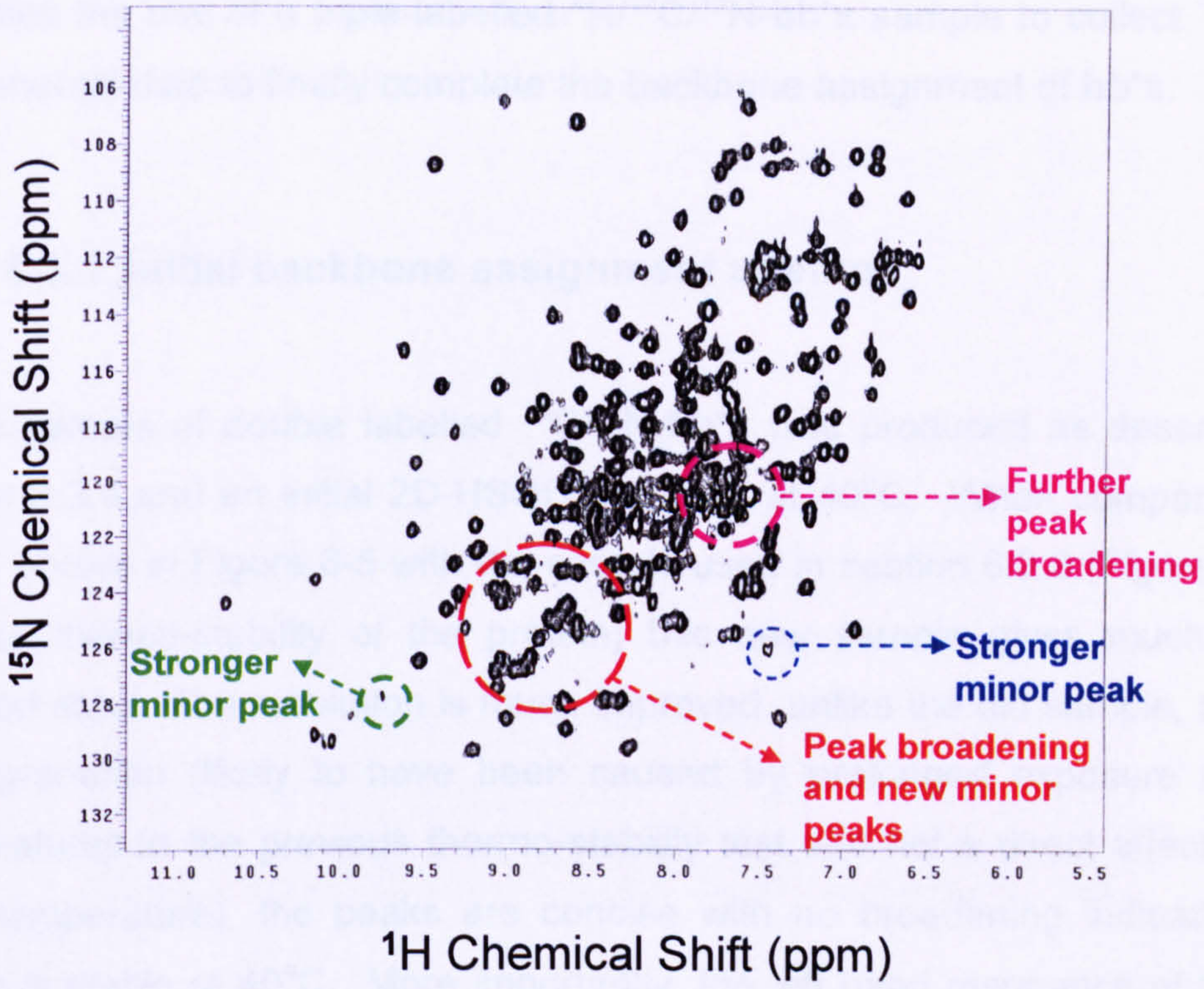

Figure 6-4. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 40°C after 22 hours.

6.3. Backbone Assignment using 40°C Triple Resonance Data

The initial triple resonance data collected on **bb'x** at 25°C, showed a lack of resolution in the HNCACB, but the thermo-stability analysis of **bb'x** at a temperature of 40°C identified a possible method of improving resolution and so enabling backbone assignment. This section begins with a description of how the triple resonance data collected on **bb'x** at 40°C was used and how the pre-assigned **b** domain HSQC was used to partially assign the backbone. Unfortunately, the dataset collected at 40°C was insufficient to assign the backbone completely. Therefore this section goes on to describe how it was essential to return to the **b'x** domain construct to obtain backbone assignments of **b'x** and cross reference these assignments to allow further assignment of **bb'x**. Yet again, full confident assignment was not possible. This section also

describes the use of a triple-labelled $^{2}H/^{13}C/^{15}N$-**bb'x** sample to collect TROSY experimental data to finally complete the backbone assignment of **bb'x**.

6.3.1. Initial backbone assignment attempt

A new sample of double labelled $^{13}C/^{15}N$-**bb'x** was produced as described in section 3.3.4 and an initial 2D-HSQC collected at 40°C. When comparing the HSQC shown in Figure 6-5 with the sample used in section 6.2.3 (Figure 6-4 to test the thermo-stability of the protein) this new sample gives much better resolved data. The resolution is much improved, unlike the old sample, there is no degradation (likely to have been caused by prolonged exposure to high temperatures in the previous thermo-stability test and not a direct affect of the 40°C temperature), the peaks are concise with no broadening indicating the protein is stable at 40°C. More importantly, the left hand resonance of the two signals from the Trp indole side chain is the more prominent and so a good indicator that the fractionation was effective at separating the two conformers.

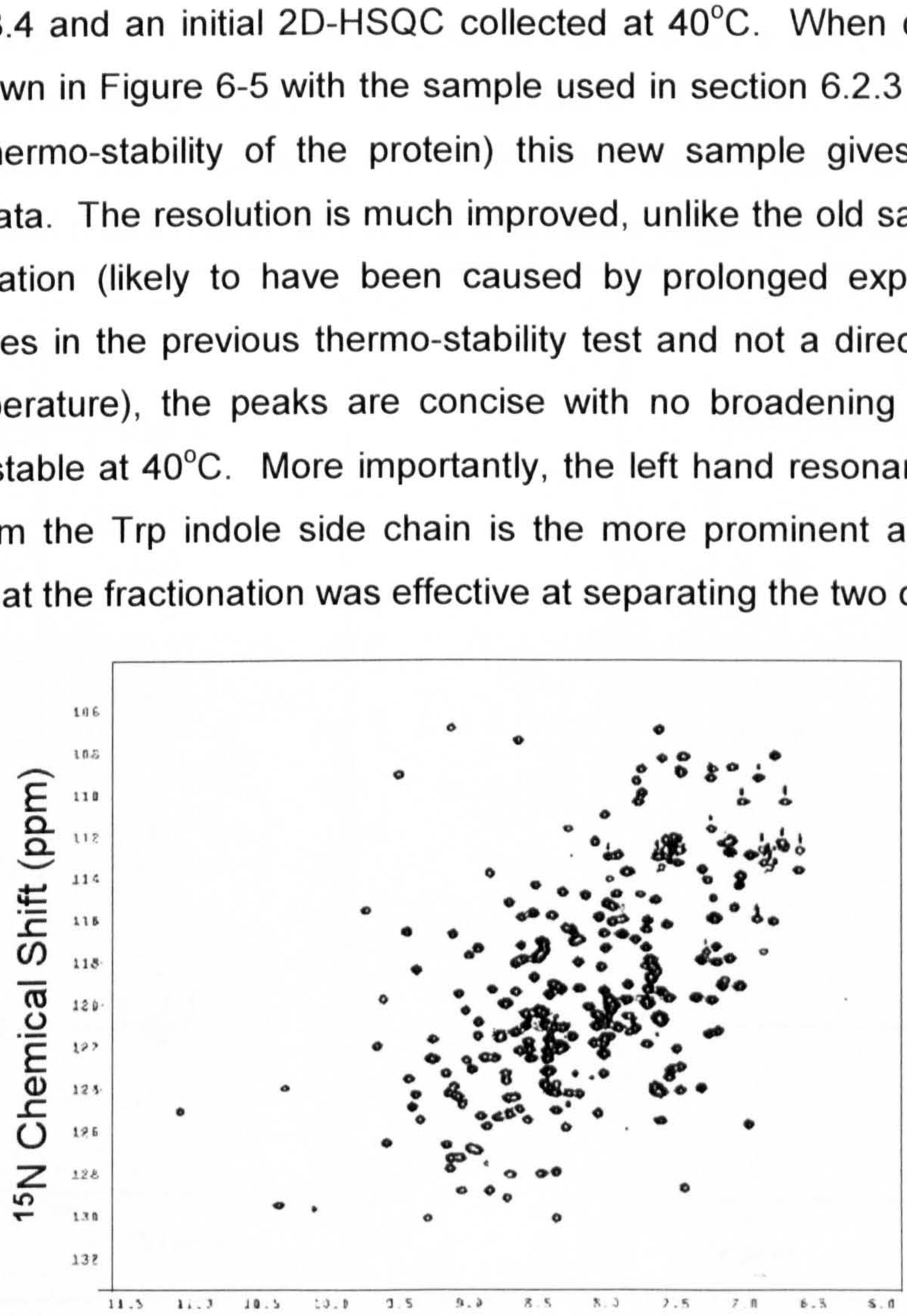

Figure 6-5. $^{15}N/^{1}H$ HSQC spectrum of a new $^{13}C/^{15}N$-bb'x domain sample at 40°C

This sample was then used to obtain triple resonance data, HNCACB and CBCA(CO)NH spectra each taking 4 days to collect data. A typical HNCACB and CBCA(CO)NH spectrum data set for a residue (Gly151) of **bb'x** at 40°C is shown is Figure 6-6. The HNCACB experiment was run first since this experiment is inherently less sensitive, which is clear from the weaker signals observed in this spectrum. The CBCA(CO)NH was run second and it is also clear that this experiment is far more sensitive as the signals in this spectrum are far stronger. The two spectra collected on the **bb'x** sample were used to make sequence specific sequential assignments. However, as a result of the weak signals in the HNCACB spectrum it was expected that assignment would be difficult and might require the use of alternative experiments and data.

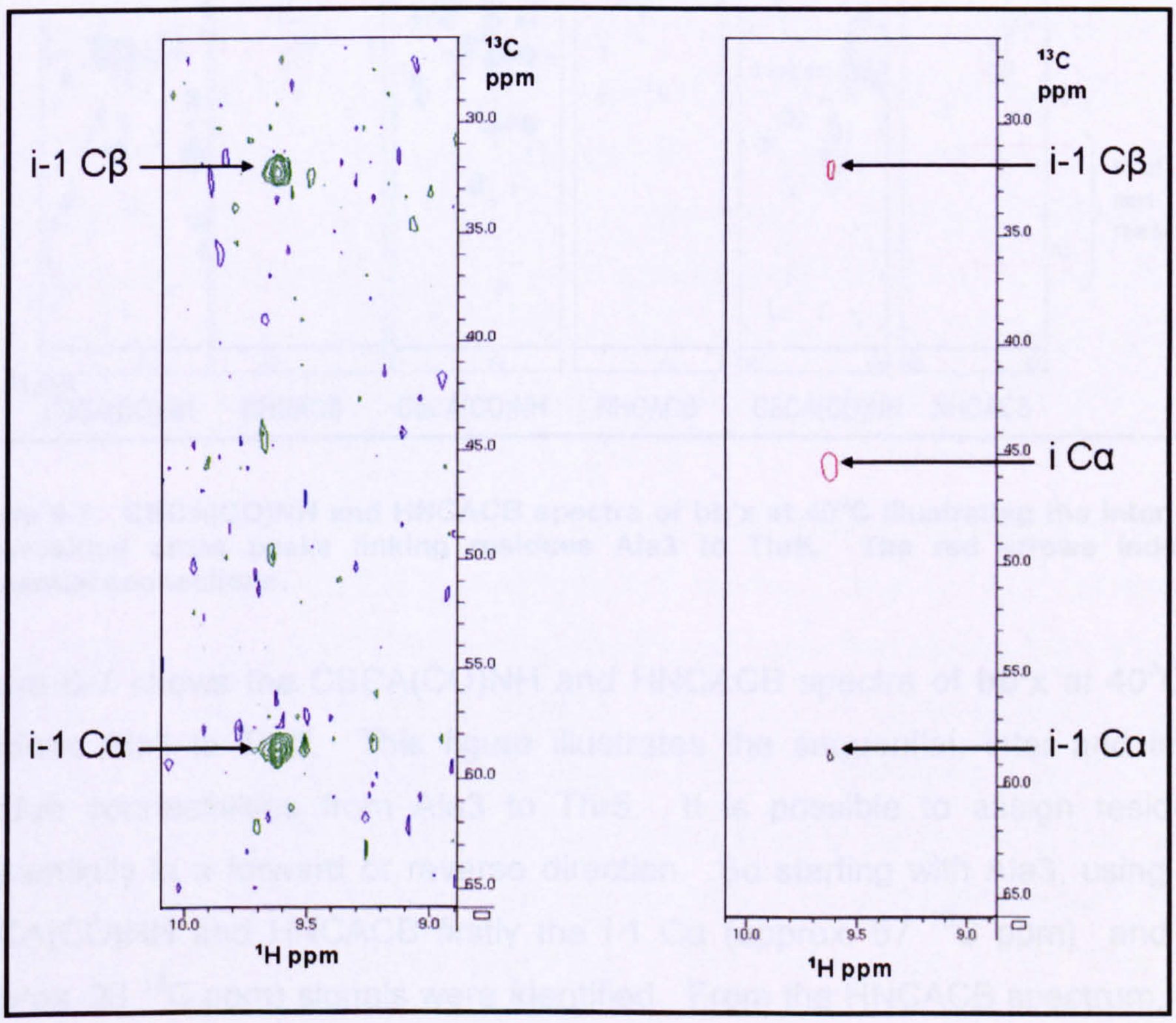

Figure 6-6. Typical ^{15}N plane (115.35 ppm) of CBCA(CO)NH (left) and HNCACB (right) spectra for what was later determined to be Gly151 of bb'x at 40°C. In the CBCA(CO)NH the i-1 Cα and Cβ, positive peaks are shown in green (Cα and Cβ) and negative peaks shown in purple. In the HNCACB show the i and i-1 peaks, where the negative plane peaks are shown in red and positive plane peaks shown in black.

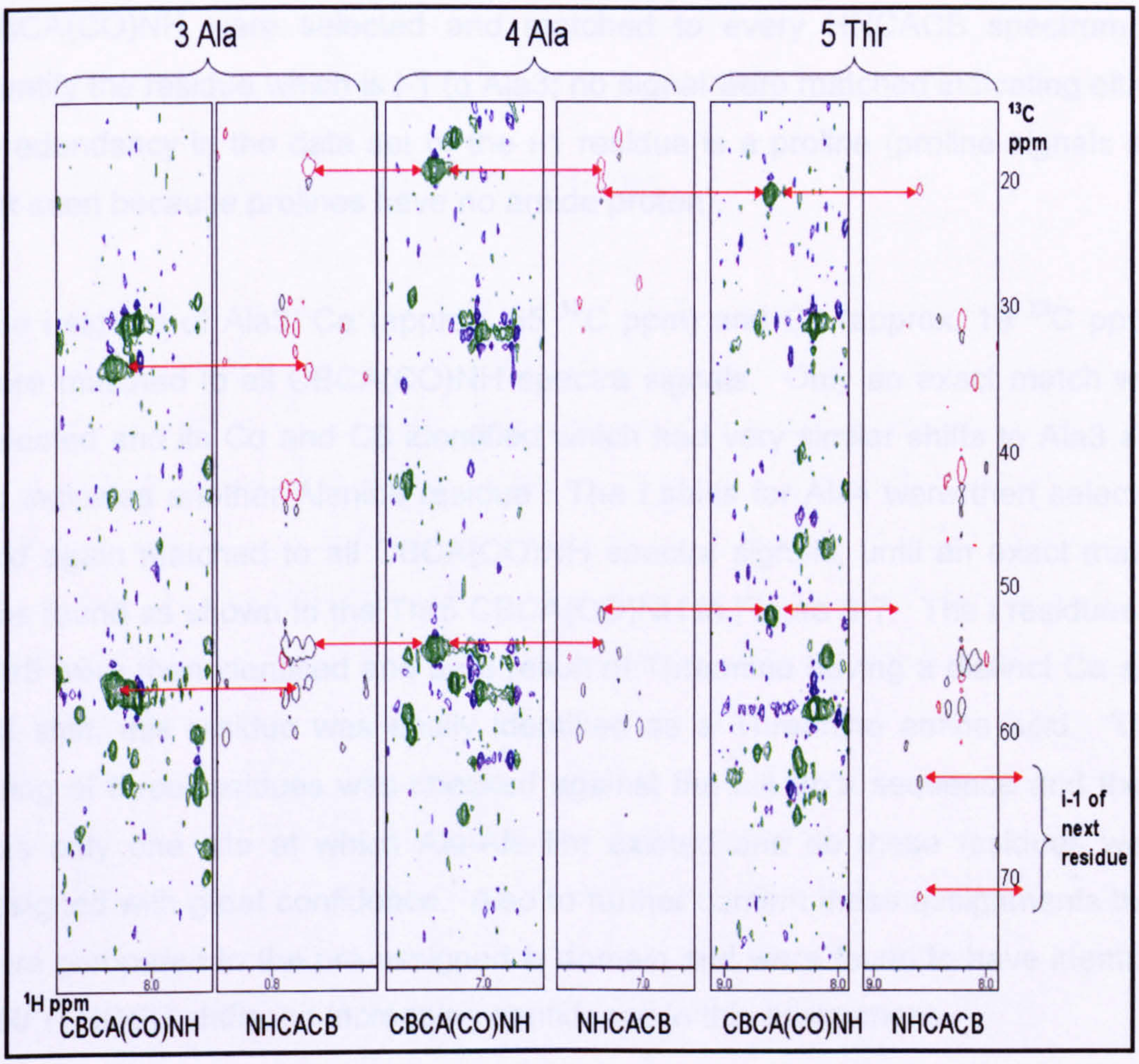

Figure 6-7. CBCA(CO)NH and HNCACB spectra of bb'x at 40°C illustrating the inter and intra-residue cross peaks linking residues Ala3 to Thr5. The red arrows indicate sequential connections.

Figure 6-7 shows the CBCA(CO)NH and HNCACB spectra of **bb'x** at 40°C of residues Ala3 to Thr5. This figure illustrates the sequential, inter and intra-residue connectivities from Ala3 to Thr5. It is possible to assign residues sequentially in a forward or reverse direction. So starting with Ala3, using the CBCA(CO)NH and HNCACB firstly the i-1 Cα (approx. 57 ^{13}C ppm) and Cβ (approx. 33 ^{13}C ppm) signals were identified. From the HNCACB spectrum, this residue was easily identified since Alanine is the only residue with a Cβ shift of approximately 20 ^{13}C ppm shift. Residues such as Alanine, Serine, Threonine residues have distinct Cα and Cβ shifts, Glycine also has a distinct single Cα shift (see Figure 2-5 for standard residue chemical shifts). The i-1 peaks in the

CBCA(CO)NH were selected and matched to every HNCACB spectrum to identify the residue which is i-1 to Ala3; no signal were matched indicating either a redundancy in the data set or the i-1 residue is a proline (proline signals are not seen because prolines have no amide proton).

The i signals of Ala3, Cα (approx. 55 ^{13}C ppm) and Cβ (approx. 19 ^{13}C ppm), were matched to all CBCA(CO)NH spectra signals. Only an exact match was selected and its Cα and Cβ identified which had very similar shifts to Ala3 and so indicates another Alanine residue. The i shifts for Ala4 were then selected and again matched to all CBCA(CO)NH spectra signals, until an exact match was found as shown in the Thr5 CBCA(CO)NH in Figure 6-7. The i residues of Thr5 were then identified and as a result of Threonine having a distinct Cα and Cβ shift, this residue was easily identified as a Threonine amino acid. This string of three residues was checked against the full **bb'x** sequence and there was only one site at which Ala-Ala-Thr existed and so these residues were assigned with great confidence. Also to further confirm these assignments they were compared to the pre-assigned **b** domain and were found to have identical ^{15}N/^{1}H HQSC shifts, so increasing confidence in this assignment.

Initially all Alanine, Serine, Threonine and Glycine residues which have distinct Cα and Cβ shifts were used as anchor points; used to start chains of assignments. Using the method described and cross referencing from the pre-assigned **b** domain, sequence specific assignments were obtained for all residues within sequential stretches Ala4 to Leu101. Residues 1 to 101 make up the **b** domain which has already been speculated to be far more stable and here almost completely assigned (with the exception two Proline residues) aided by the pre-assigned **b** domain discussed in the next section 6.3.2. This is in contrast to the **b'x** domain region in **bb'x**, where sequence specific assignments were only obtained for sequential stretches from Thr108 to Ala112, Ala145 to Ile158, and Thr202 to His211; this relates to only 34% of the **b'x** region in **bb'x**.

Clearly the **b'x** domain is more conformationally flexible, resulting in low spectral resolution and so causing difficulty in sequence-specific assignments.

6.3.2. b domain cross assignment

The **b** domain was fully assigned by Dr.K.Wallis (University of Warwick). This data set was used to cross reference the assignments made in **b** domain on to the **bb'x** HSQC at 40°C. This was done to reduce ambiguity when assigning **bb'x**. This cross assignment was done initially by overlaying the **bb'x** HSQC collected at 40°C with the **b** domain spectrrum collected at 25°C as shown in Figure 6-8. The temperature difference will inherently result in some resonance shifts as highlighted in the yellow circle of the expanded HSQC in Figure 6-8, but many of the resonances overlaid exactly as highlighted with the green circle in the expanded HSQC in Figure 6-8. Since every residue was assigned in the **b** domain HSQC this information was cross referenced and checked if the assignments made in the **bb'x** data set have equivalent ^{1}H and ^{15}C HSQC shifts. Once located in the HSQC, the **bb'x** CBCA(CO)NH and HNCACB were re-analysed with lower contour levels to identify the Cα and Cβ resonances. This method reduces ambiguity and enabled all the **b** domain residues to be assigned in the **bb'x** data set with the obvious exceptions of Proline.

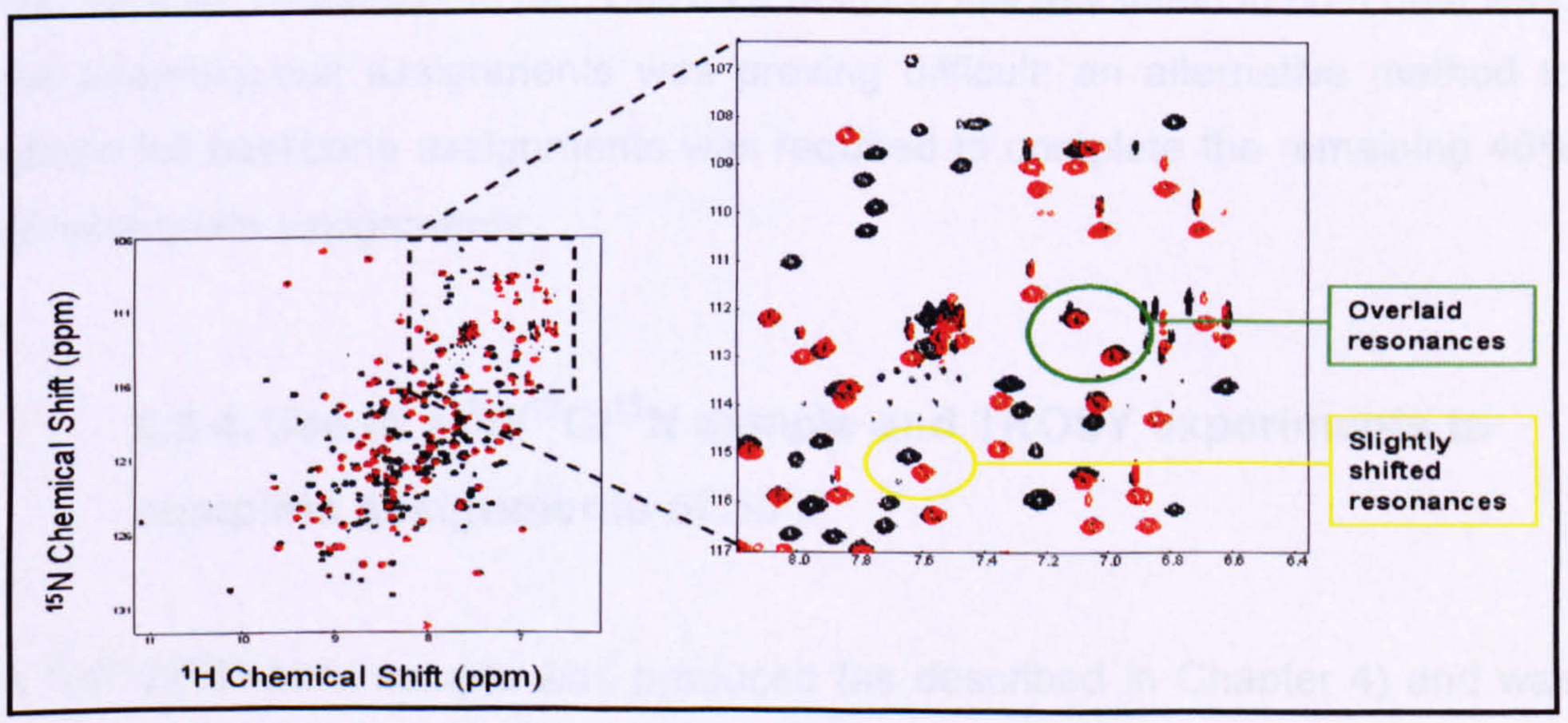

Figure 6-8. $^{15}N/^{1}H$ HSQC spectrum of $^{13}C/^{15}N$-bb'x domain at 40°C (black peaks) overlaid with ^{15}N-b domain (red peaks). Right HSQC – expanded image of overlaid HSQC spectrum (dashed box in the left spectrum).

6.3.3. b'x backbone assignment to assist the assignment of the bb'x backbone

Using the 40°C **bb'x** data, only 34% of the **b'x** domain was assigned. So to obtain further assignments, attention was directed at the **b'x** domain. The 2D-HSQC described in section 5.3.2, was produced using a ^{15}N-labelled **b'x** sample, as described in section 3.4.1. The 2D-HSQC spectral quality was high and so in this situation whereby assignment of **bb'x** has emerged to be difficult, a step back to **b'x** could solve these difficulties. A double-labelled **b'x** sample was produced as described in section 3.4.2 and the 2D-HSQC of this sample was comparable in resolution to that obtained in section 5.3.2 with the ^{15}N-labelled **b'x**. The **b'x** backbone assignments were made as described in section 6.3.1 for **bb'x** using the same CBCA(CO)NH and HNCACB triple resonance experimental data collected using the ^{13}C/^{15}N **b'x** sample. Using these data sets, 82 of the 136 (60%) **b'x** domain residues were assigned, with the aid of the **b'x** segment assignments made using **bb'x** in section 6.3.1. Therefore, an extra 26% (60% **b'x** assignments - 34% **b'x** assignments already made in **bb'x** and described in section 6.3.1) could be cross referenced to the **bb'x** data set and assigned. Hence 60 % of the **b'x** region in **bb'x** was assigned using the two data set described so far. Still as a result of low resolution in both data sets, full unambiguous assignments was proving difficult; an alternative method to obtain full backbone assignments was required to complete the remaining 40% of incomplete assignments.

6.3.4. Use of a ^{2}H/^{13}C/^{15}N sample and TROSY experiments to complete assignments of bb'x

A ^{2}H/^{13}C/^{15}N **bb'x** sample was produced (as described in Chapter 4) and was used to obtain HNCA and HN(CO)CA 3D-data sets; methodology described in section 1.4.4. The advantage of the deuterated sample is that because

deuterons have an approximately 6.5 times lower gyromagnetic ratio than that of a proton, a large amount of the fluctuating magnetic fields can be removed; since fluctuating magnetic fields reduce spectral signals strength. So the deuterated sample should provide data with greater resolution. This is illustrated in Figure 6-9, whereby the 2D-HSQC has greater resolution, the peaks especially in the centre of the spectrum are far more resolved and the amount of overlap in this region is greatly reduced when compared with the 2D-HSQC collected with the double-labelled **bb'x** sample shown in Figure 6-5.

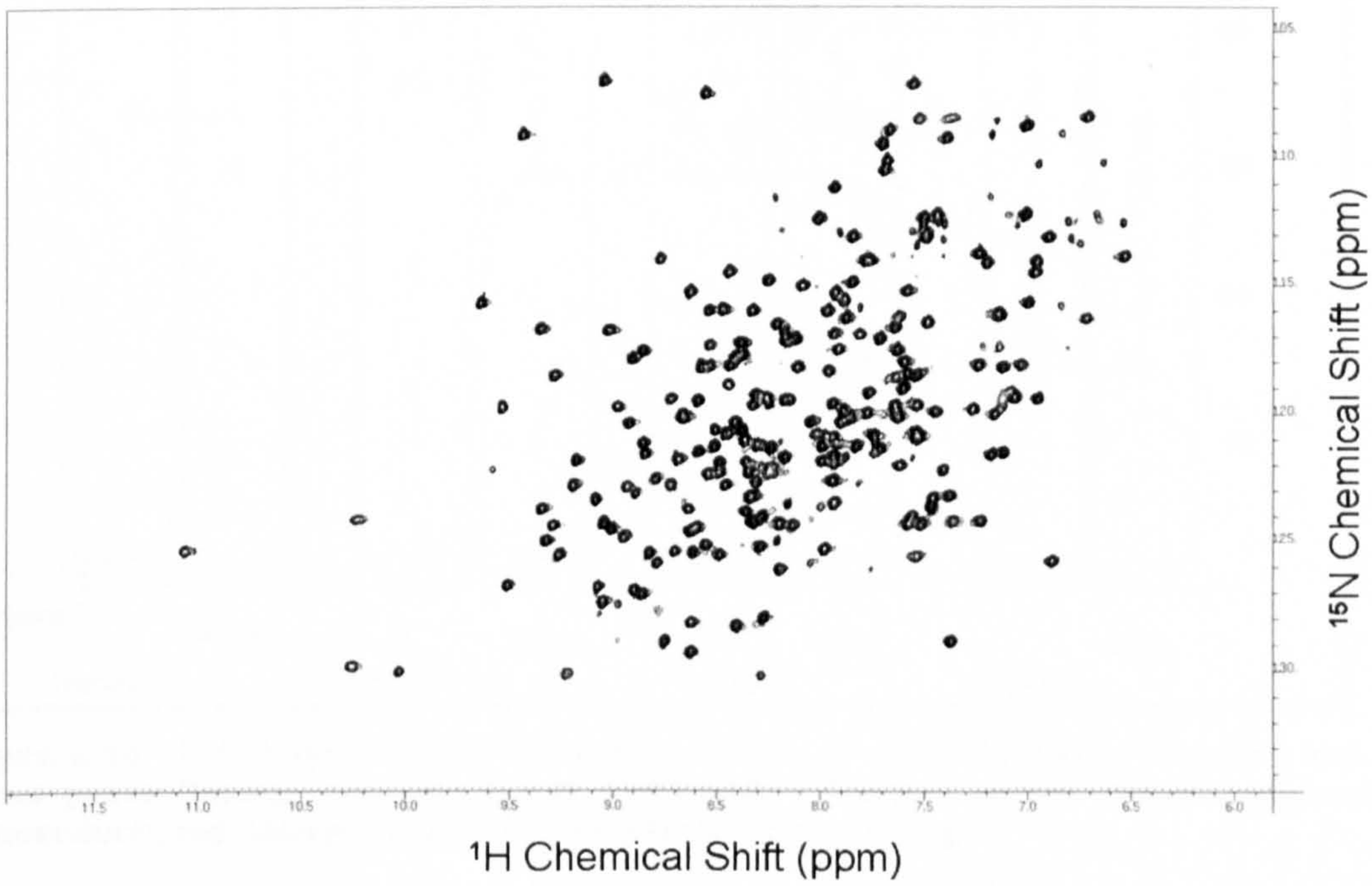

Figure 6-9. $^{15}N/^{1}H$ HSQC spectrum of $^{2}H/^{13}C/^{15}N$-bb'x domain at 40°C.

As expected the HNCA and HN(CO)CA experiments were seen to be far more sensitive, with greater resolution as shown in Figure 6-10 in comparison with the CBCANH and CBCA(CO)NH data shown in Figure 6-6. These three dimensional experiments have a proton axis, a ^{15}N axis and a ^{13}C axis. Each residue i peak will give a (H_N(i), N(i), Cα (i)) and (H_N(i), N(i), Cα (i-1)) peaks while the complementary HN(CO)CA will give only the (H_N(i), N(i), Cα(i-1)). In a similar method as described in section 6.3.1 adjacent residues in the **bb'x**

sequence can be linked using in this case only the Cα peaks and so sequentially assigned.

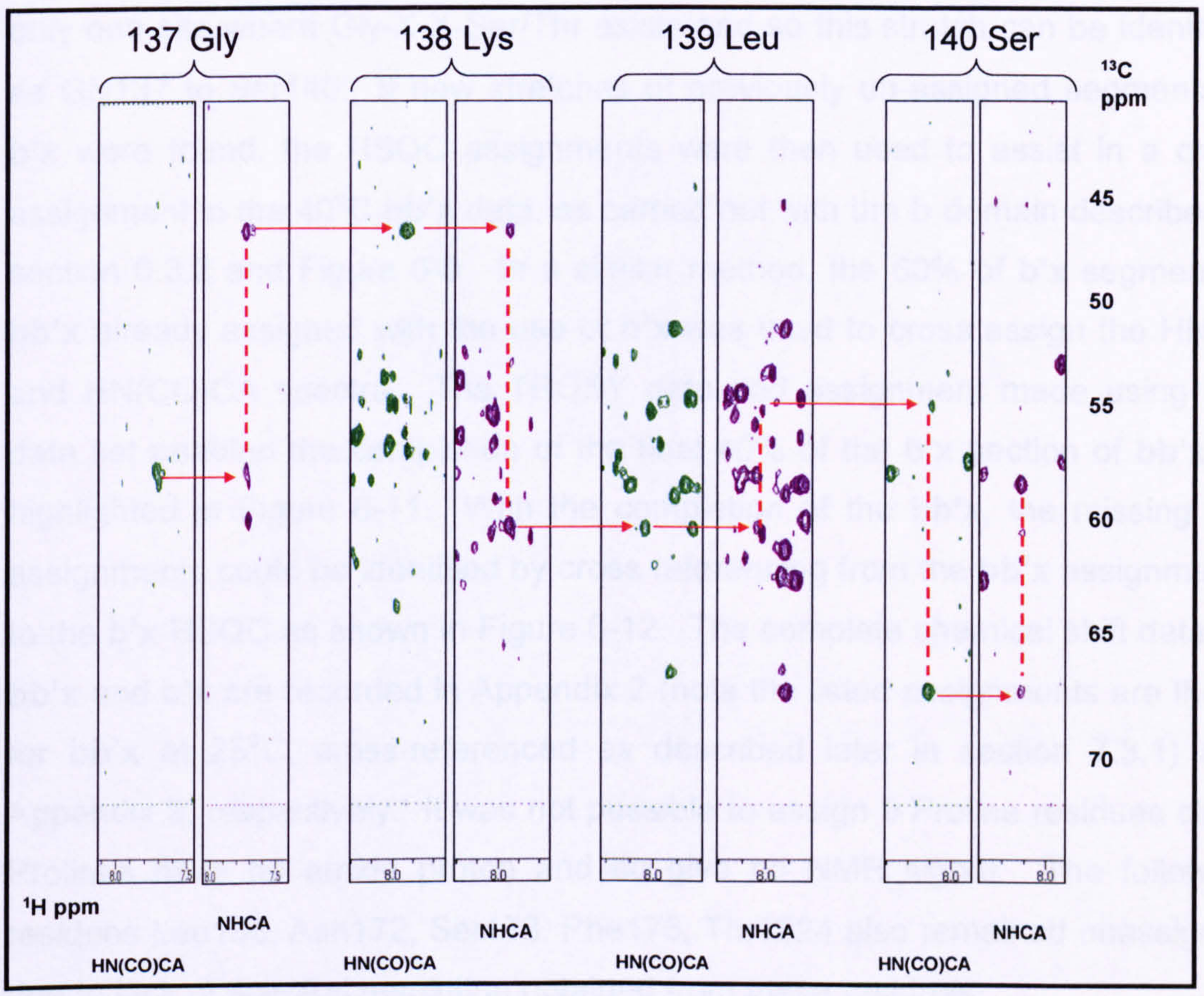

Figure 6-10. HNCA and HN(CO)CA spectra of bb'x at 40°C illustrating the Cα i and i-1 cross peaks linking residues Gly137 to Ser140. The red arrows indicate sequential connections, red dashed lines indicated inter-spectrum linkages.

Several residues were initially identified as anchor points since they have distinct Cα shifts, such as Glycine, Alanine, Serine and Threonine. As shown in Figure 6-10 residue 137 was selected and immediately identified as a Glycine residue from the distinct Cα shift of approximately 45ppm (^{13}C dimension). The associated i Cα peak is then identified since it lies directly on the same ^{1}H dimension plane in the NHCA spectrum (some ambiguity may exist here, if several peaks are visible and so they needed to be checked by scrolling through the ^{15}N plane and identifying which residues were also on the same ^{15}N plane). This i peak was then interrogated against all HN(CO)CA peaks until a exact

match was found in another residue. This process was repeated until another distinct peak was found and in this case, in Figure 6-10, the final spectrum on the right is likely to be a Serine or Threonine and from the sequence there is only one site where Gly-X-X-Ser/Thr exists and so this stretch can be identified as Gly137 to Ser140. If new stretches of previously un-assigned segments of **b'x** were found, the HSQC assignments were then used to assist in a cross assignment in the 40°C **bb'x** data, as carried out with the **b** domain described in section 6.3.2 and Figure 6-8. In a similar method, the 60% of **b'x** segment of **bb'x** already assigned with the use of **b'x** was used to cross assign the HNCA and HN(CO)CA spectra. The TROSY data and assignment made using this data set enabled the completion of the final 40% of the **b'x** section of **bb'x** as highlighted in Figure 6-11. With the completion of the **bb'x**, the missing **b'x** assignments could be identified by cross referencing from the **bb'x** assignments to the **b'x** HSQC as shown in Figure 6-12. The complete chemical shift data for **bb'x** and **b'x** are recorded in Appendix 2 (note the listed assignments are those for **bb'x** at 25°C, cross-referenced as described later in section 7.3.1) and Appendix 3, respectively. It was not possible to assign 9 Proline residues since Prolines have no amide proton and so give no NMR signal. The following residues Leu156, Asn172, Ser173, Phe176, Thr1224 also remained unassigned due to lack of spectral resolution obtained from these residues.

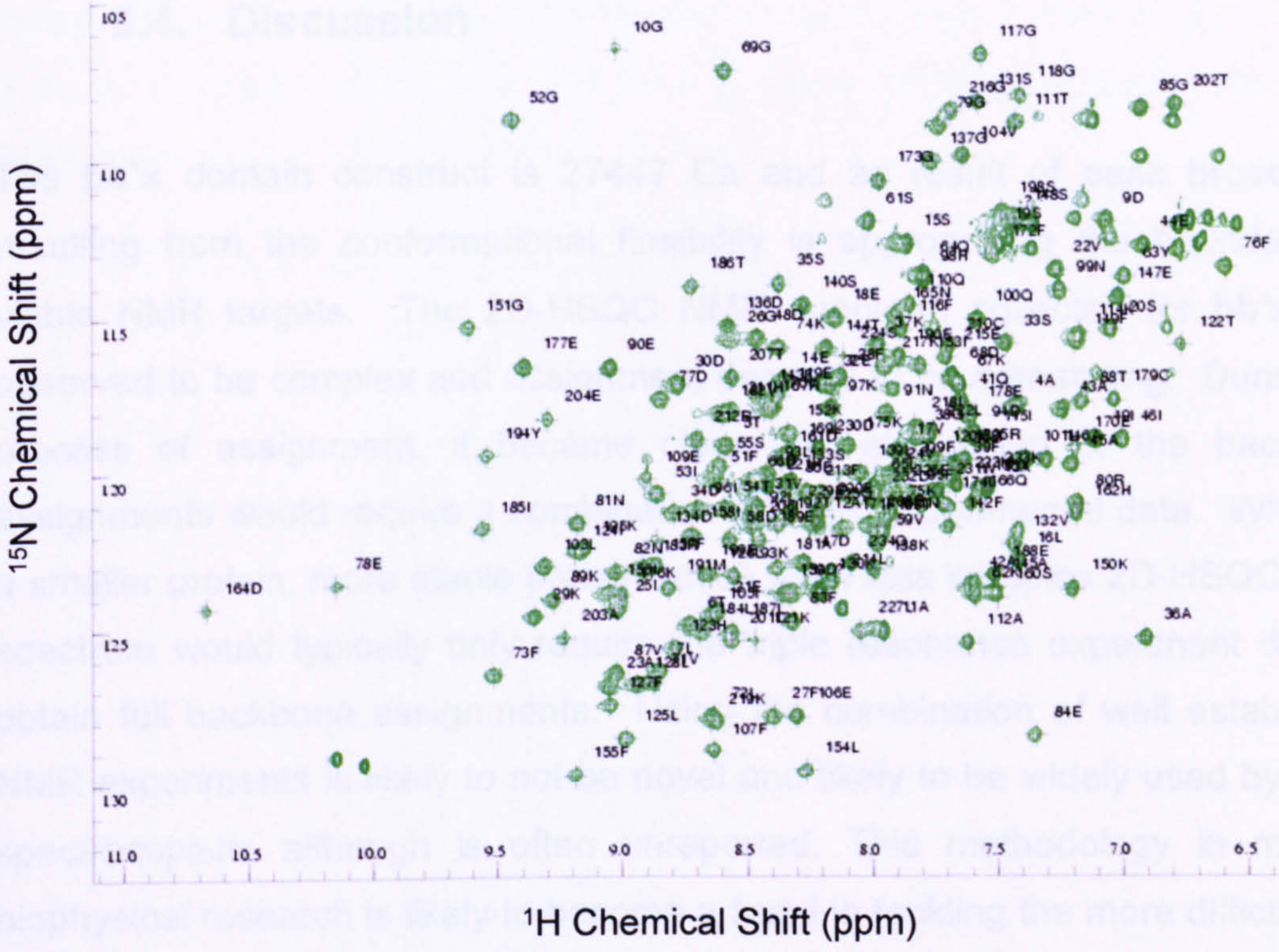

Figure 6-11. Fully assigned bb'x HSQC at 40°C. Labelled with construct numbers and single letter amino acid code.

Figure 6-12. Fully assigned b'x HSQC at 25°C. Labelled with construct numbers and single letter amino acid code.

6.4. Discussion

The **bb'x** domain construct is 27447 Da and as result of peak broadening resulting from the conformational flexibility is approaching the boundary for viable NMR targets. The 2D-HSQC NMR spectrum collected for **bb'x** was observed to be complex and assignment deemed to be demanding. During the process of assignment, it became clear that elucidation of the backbone assignments would require a combination of NMR experimental data. Whereas a smaller protein, more stable protein which gave less complex 2D-HSQC NMR spectrum would typically only require the triple resonance experiment data to obtain full backbone assignments. Using the combination of well established NMR experiments is likely to not be novel and likely to be widely used by NMR spectrocopists; although is often unreported. This methodology in modern biophysical research is likely to become a trend in tackling the more difficult and larger protein targets in research.

As shown below in Figure 6-13, the demanding task of assignment is summarised. The process begins by exhausting the triple resonance data collected on **bb'x**, to assign as much as possible with this dataset, iteratively. Once this process was exhausted, the second step was to cross reference the **b** domain assignments (assigned by Dr. K. Wallis) on to the **bb'x**, leading to full assignment of the **b** domain in the **bb'x** spectrum. To complete the **bb'x** assignments, the **b'x** was assigned as much as possible and then cross-referenced on to **bb'x**. This method was advantageous for both datasets, as redundant data in each set of experiments could be complemented and completed in the other; so aid the assignment of each protein (**b'x** and **bb'x**). The final step to enable completion required TROSY data collected using a triple-labelled sample, the preparation of which is discussed in the next chapter.

This elaborate method was not initially expected, but proved very effective in achieving the goal of full backbone assignment of a large and demanding

protein molecule. The fruit of this labour is that it allows further analysis of the domain construct and paves the way to allow probing dynamic and more detailed molecular NMR analysis.

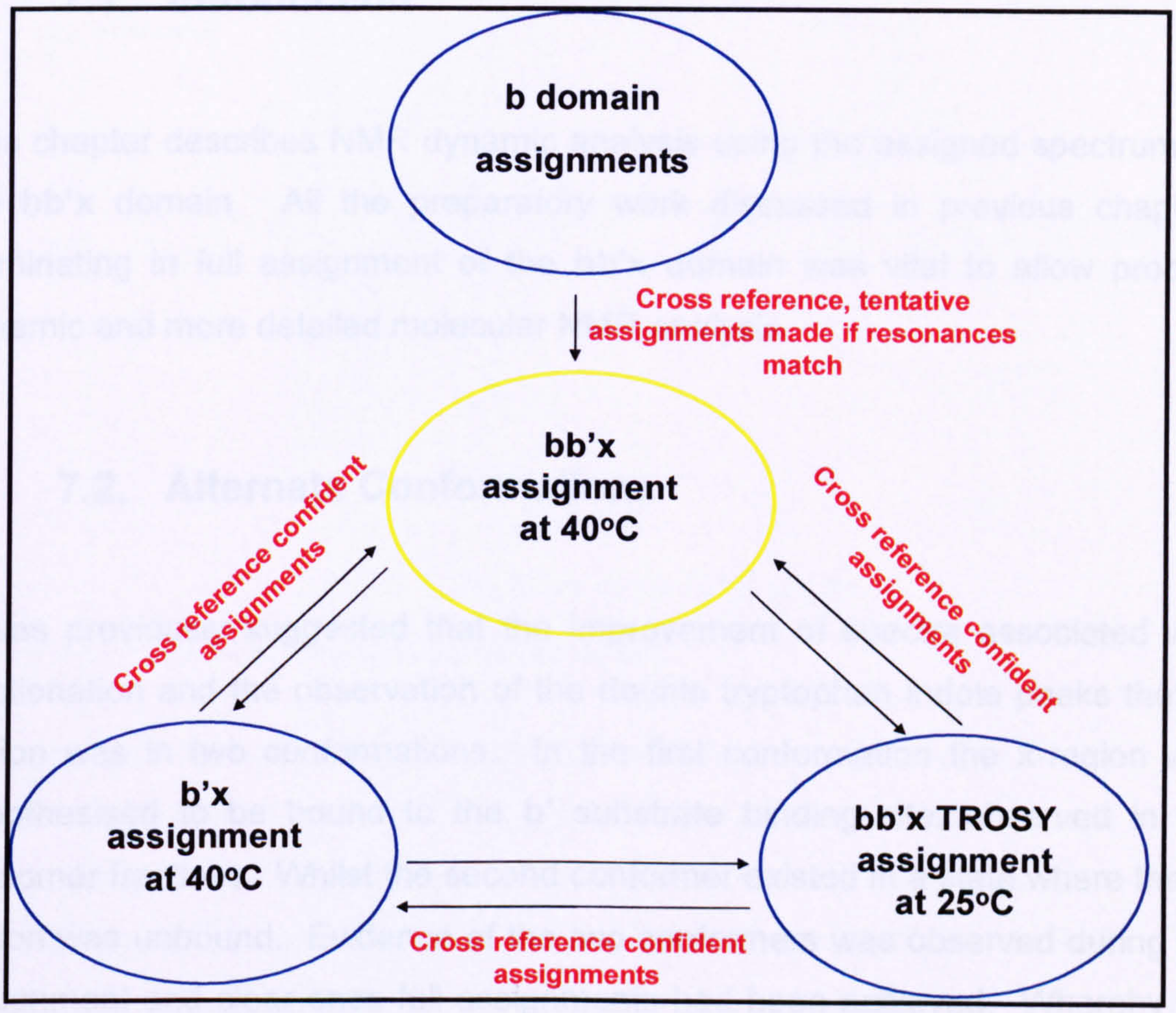

Figure 6-13. Schematic diagram of the assignment method using the different experimental NMR data collected.

Chapter 7. Protein Conformation and Dynamics

7.1. Introduction

This chapter describes NMR dynamic analysis using the assigned spectrum for the **bb'x** domain. All the preparatory work discussed in previous chapters culminating in full assignment of the **bb'x** domain was vital to allow probing dynamic and more detailed molecular NMR analysis.

7.2. Alternate Conformations

It was previously suggested that the improvement of spectra associated with fractionation and the observation of the double tryptophan indole peaks that **x**-region was in two conformations. In the first conformation the **x**-region was hypothesised to be bound to the **b'** substrate binding site, observed in the monomer fractions. Whilst the second conformer existed in a state where the **x**-region was unbound. Evidence of the two conformers was observed during the assignment and clear once full assignments had been achieved. Whereby the resonances resulting from the **x**-region in the 2D-HSQC were significantly stronger and better resolved than any other peaks in the HSQC, as shown in Figure 7-1. The residues from the **x**-region such as Glu226, Trp231, Met223 (construct numbers; mature PDI numbers are Glu342, Trp347, Met339) appear much stronger with many more contours than any other peaks observed in the spectrum. This would infer the **x**-region was in a much more stable/rigid conformation.

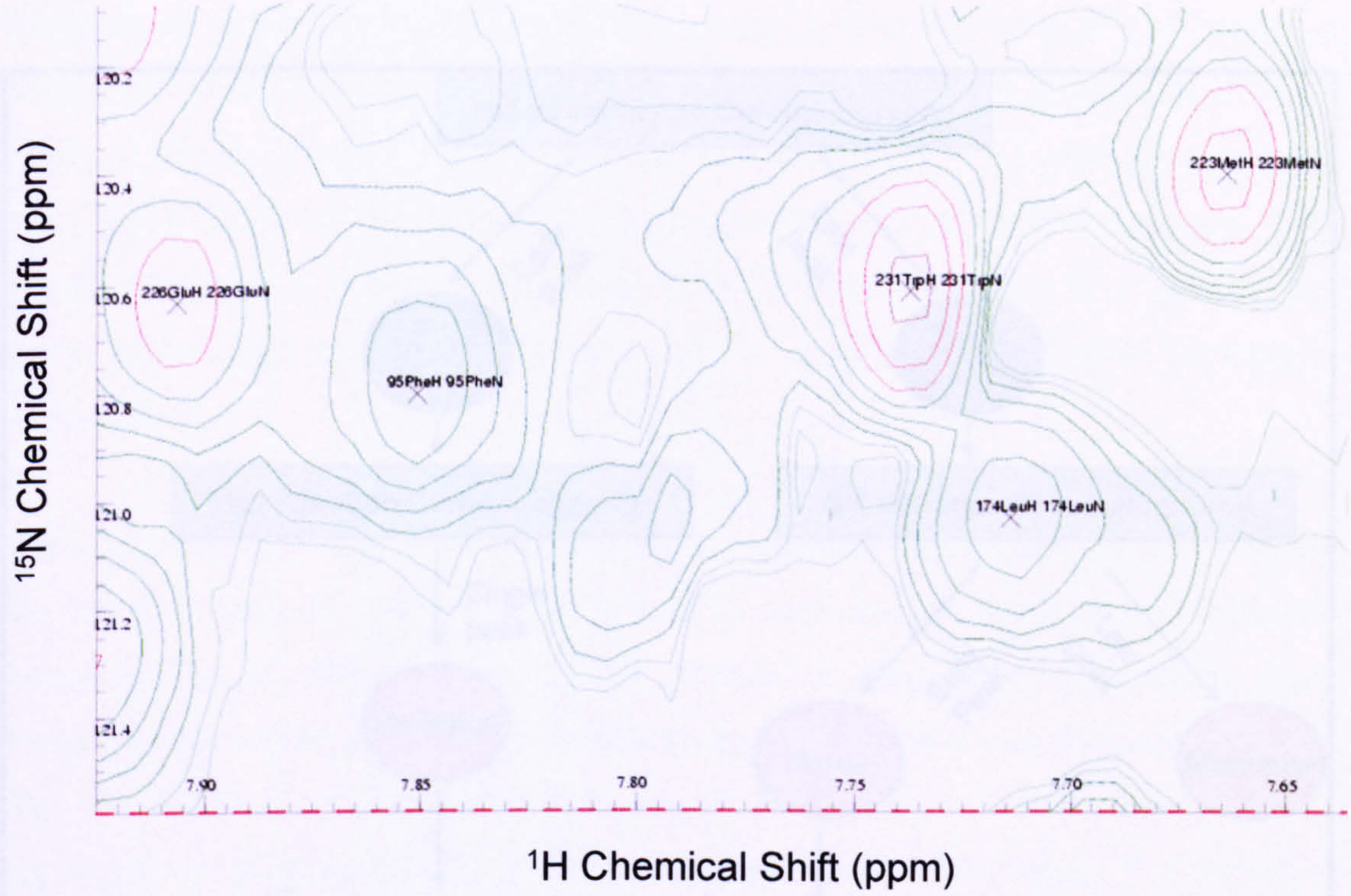

Figure 7-1. Zoomed in image of bb'x HSQC at 40°C.

The species observed during the preparation and characterisation of the **bb'x** protein samples were defined in Figure 3-22, based on observation made during the purification and fractionation. With these new observations from the NMR data, these definitions of species require redefining, as shown in Figure 7-2. The Pool 1 monomer fractions, were initially defined as containing a monomer species, but in light of the NMR data, this species must be redefined as sample containing predominately a 'capped monomer' species where the **x**-region is speculated to be bound to the **b'** substrate binding site. Whilst the Pool 2 dimer sample, initially defined as containing primarily a dimer species, but must be redefined as sample containing some of the dimer species and an 'un-capped monomer' species where the **x**-region is not bound to the **b'**domain. It was still unclear as to the proportions of the dimer and uncapped monomer species which exist in the Pool 2 dimer samples.

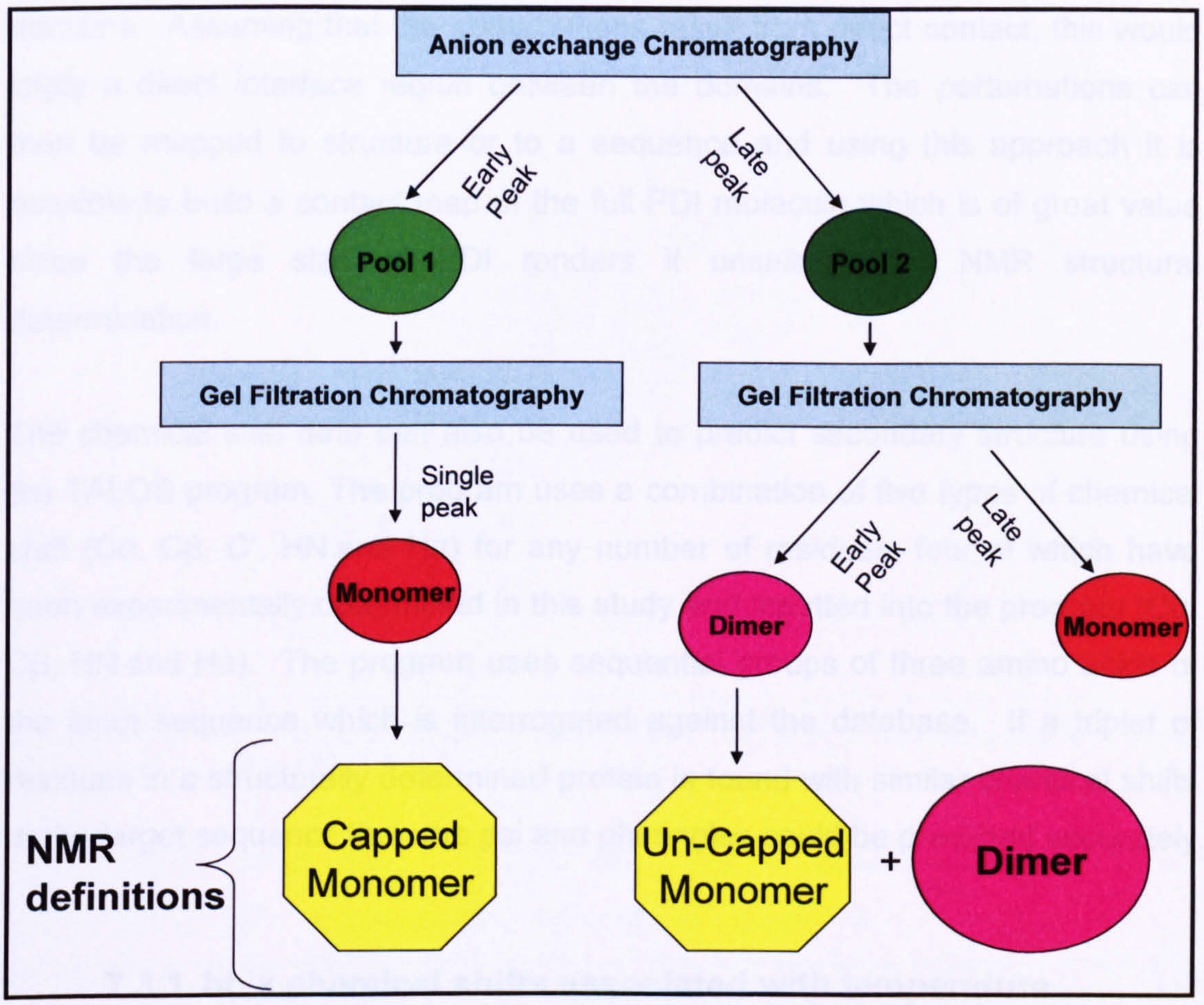

Figure 7-2. Diagram showing the redefinition of species observed through various stages of characterisation and analysis.

7.3. Chemical Shift Analysis

One type of molecular analysis involves using the chemical shift data which can be employed to investigate the effect of temperature change on specific residues in the **bb'x** domain construct. Another involves comparative analysis of spectra of different domain constructs to define inter-domain interactions/interfaces. An assigned HSQC spectrum, for example the **bb'x** construct, can be compared to another assigned HSQC, such as that of the **b** domain, and any perturbations of the **b** domain resonances in **bb'x** recorded. These perturbations that arise would highlight the residues involved in direct

interactions or long distance structural changes arising from addition of adjacent domains. Assuming that the perturbations result from direct contact, this would imply a direct interface region between the domains. The perturbations can then be mapped to structure or to a sequence and using this approach it is possible to build a contact map of the full PDI molecule which is of great value since the large size of PDI renders it unsuitable for NMR structural determination.

The chemical shift data can also be used to predict secondary structure using the TALOS program. The program uses a combination of five types of chemical shift (Cα, Cβ, C', HN and Hα) for any number of residues; four of which have been experimentally determined in this study and inputted into the program (Cα, Cβ, HN and Hα). The program uses sequential groups of three amino acids of the input sequence which is interrogated against the database. If a triplet of residues in a structurally determined protein is found with similar chemical shifts to the target sequence then the psi and phi angles could be predicted accurately.

7.3.1. bb'x chemical shifts associated with temperature

During the process of optimising NMR spectral resolution by increasing the temperature at which data was collected (described in Chapter 6), HSQC spectra were collected at 25, 35 and 40°C. This provided an opportunity to investigate how the increase in temperature affected residues determined by chemical shift. It was determined that assignment of the **bb'x** domain could be possible using data collected at 40°C and assignments were made using the HSQC and triple resonance data collected at 40°C. Once the assignment was complete, it was possible to cross-assign simply by overlaying the spectra as shown in Figure 7-3, transferring the assignments from the 40°C HSQC to the 35°C data and then finally to the 25°C data. The majority of peaks had sufficiently similar chemical shifts at each temperature to allow unambiguous

cross-assignment; occasionally overlapping did hinder assignment but upon close analysis all peaks were assigned.

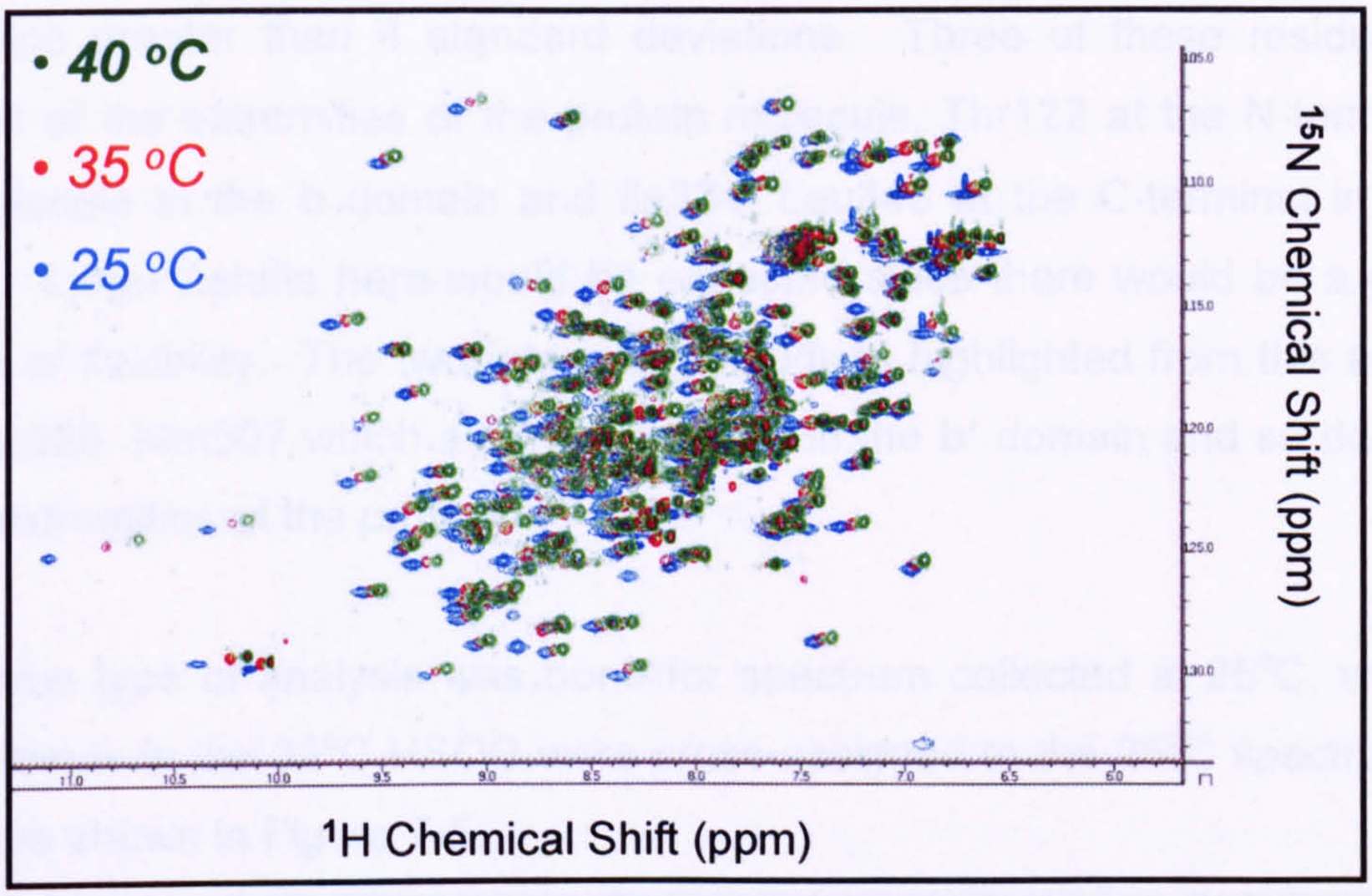

Figure 7-3. $^{15}N/^{1}H$ HSQC overlaid spectrum of a $^{13}C/^{15}N$-bb'x domain sample at 40°C (green), 35°C (red) and 25°C (blue).

The chemical shifts from the 40°C assignments were compared to those transferred to 35°C and are presented below in Figure 7-4.

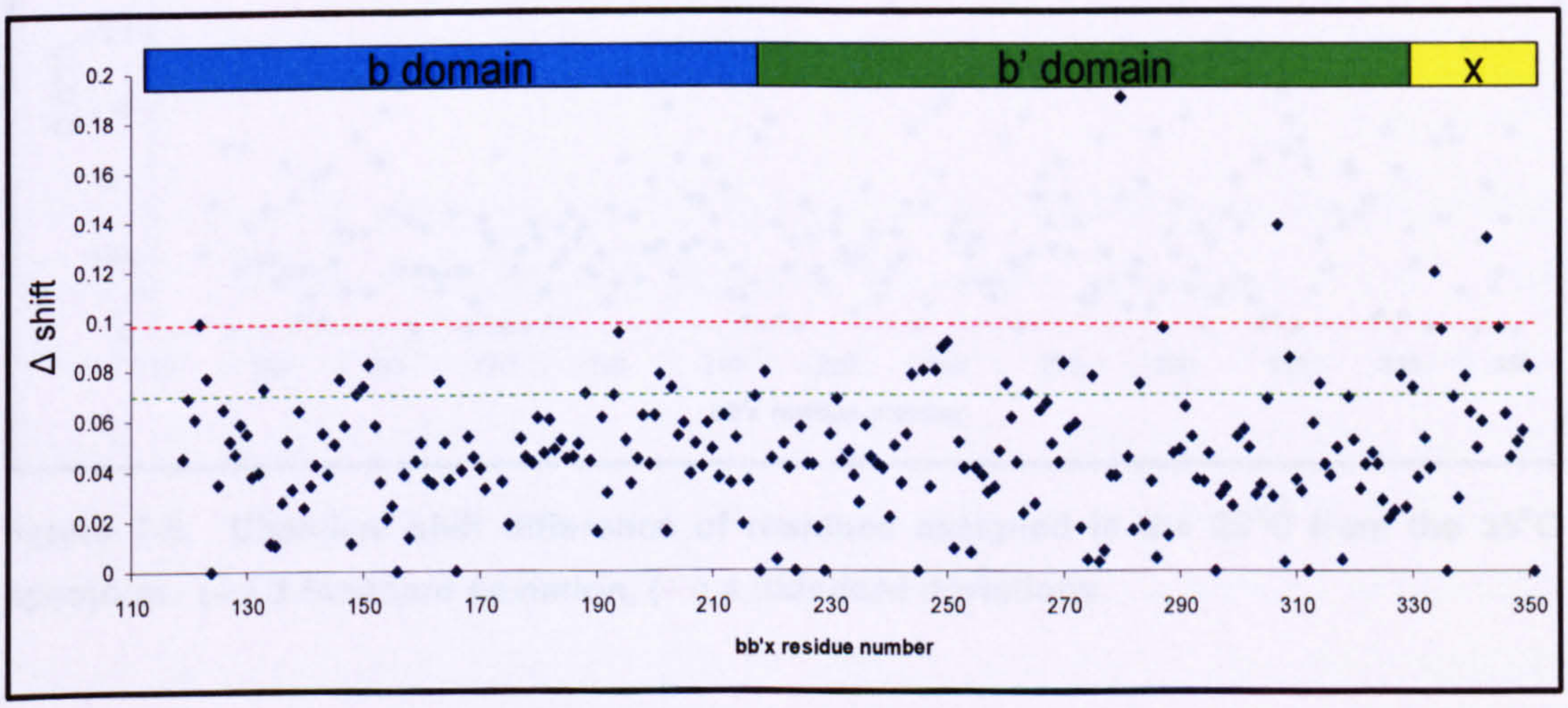

Figure 7-4. Chemical shift difference of residues assigned in the 35°C from the 40°C spectrum. (---) 3 Standard deviation, (---) 4 Standard deviations.

From Figure 7-4 it is clear there are a range of Δshifts which occur as a result of the decrease in temperature, the most significant are those which shift with a difference greater than 4 standard deviations. Three of these residues are situated at the extremities of the protein molecule, Thr122 at the N-terminal of the molecule in the **b** domain and Ile334, Leu343 at the C-terminal in the **x**-region. Large Δshifts here would be expected since there would be a greater degree of flexibility. The two interesting residues highlighted from this analysis are Asp280, Met307 which are both situated in the **b'** domain and so do not lie at the extremities of the protein.

The same type of analysis was done for spectrum collected at 25°C, whereby assignments in the 35°C HSQC were cross-assigned to the 25°C spectrum; the shifts are shown in Figure 7-5.

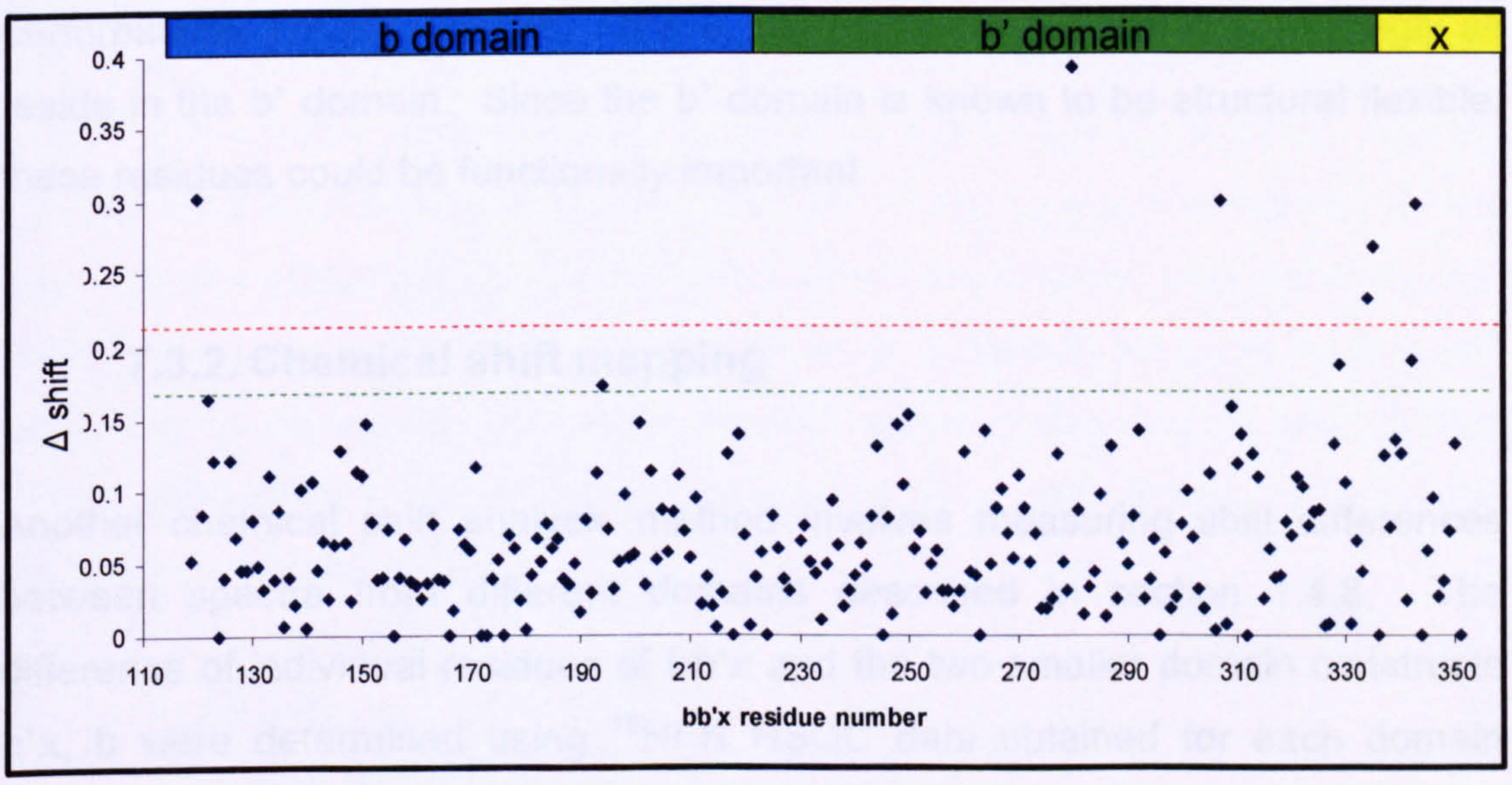

Figure 7-5. Chemical shift difference of residues assigned in the 25°C from the 35°C spectrum. (---) 3 Standard deviation, (---) 4 Standard deviations.

The same residues Ile334, Leu343 and Thr122 from the construct's extremities again produced the highest shifts in excess of 4 standard deviations, along with Asp280, Met307 as identified from the 35°C shifts. This supports the hypothesis by Dr. L. Ruddock, whose research also alluded to the importance of the L343

residue, which featured in the research to structurally determine the **b'x** domain (see Appendix 1). Other residues with shifts in excess of 4 standard deviations were Lys335 and Ala120. Also an interesting residue with a shift above 3 standard deviations is Glu342. This data suggests as expected, that the regions at the extremities of the protein are the most flexible but also two residues in the **b'** domain are also very flexible, these are Asp280, Met307. Also apparent from Figure 7-5 is that again there are a higher number of large shifts in the **b'x** region than the **b** domain. The average shifts for the **b'x** region is 0.078ppm, whereas the **b** domain averages 0.066ppm.

Alternatively the reverse analysis is also possible, whereby residues which are structurally important will have no or very low perturbations. The following residues feature in both temperature change experiments in having the very low perturbations Thr308, Glu286, Thr318, Ile274 and Ile221 which surprisingly all reside in the **b'** domain. Since the **b'** domain is known to be structural flexible, these residues could be functionally important.

7.3.2. Chemical shift mapping

Another chemical shift analysis method involves measuring shift differences between spectra from different domains described in section 1.4.8. The difference of individual residues of **bb'x** and the two smaller domain constructs **b'x**, **b** were determined using $^{15}N/^{1}H$ HSQC data obtained for each domain construct under identical conditions. The fully assigned **bb'x** HSQC is compared to the assigned **b** domain HSQC and any perturbations would infer the domain interface between the **b** domain and **b'x**. The perturbations observed between the spectra would result from the influence of the adjacent domain and since the **b** domain has been structurally determined, the interface residues can be mapped on to the structure.

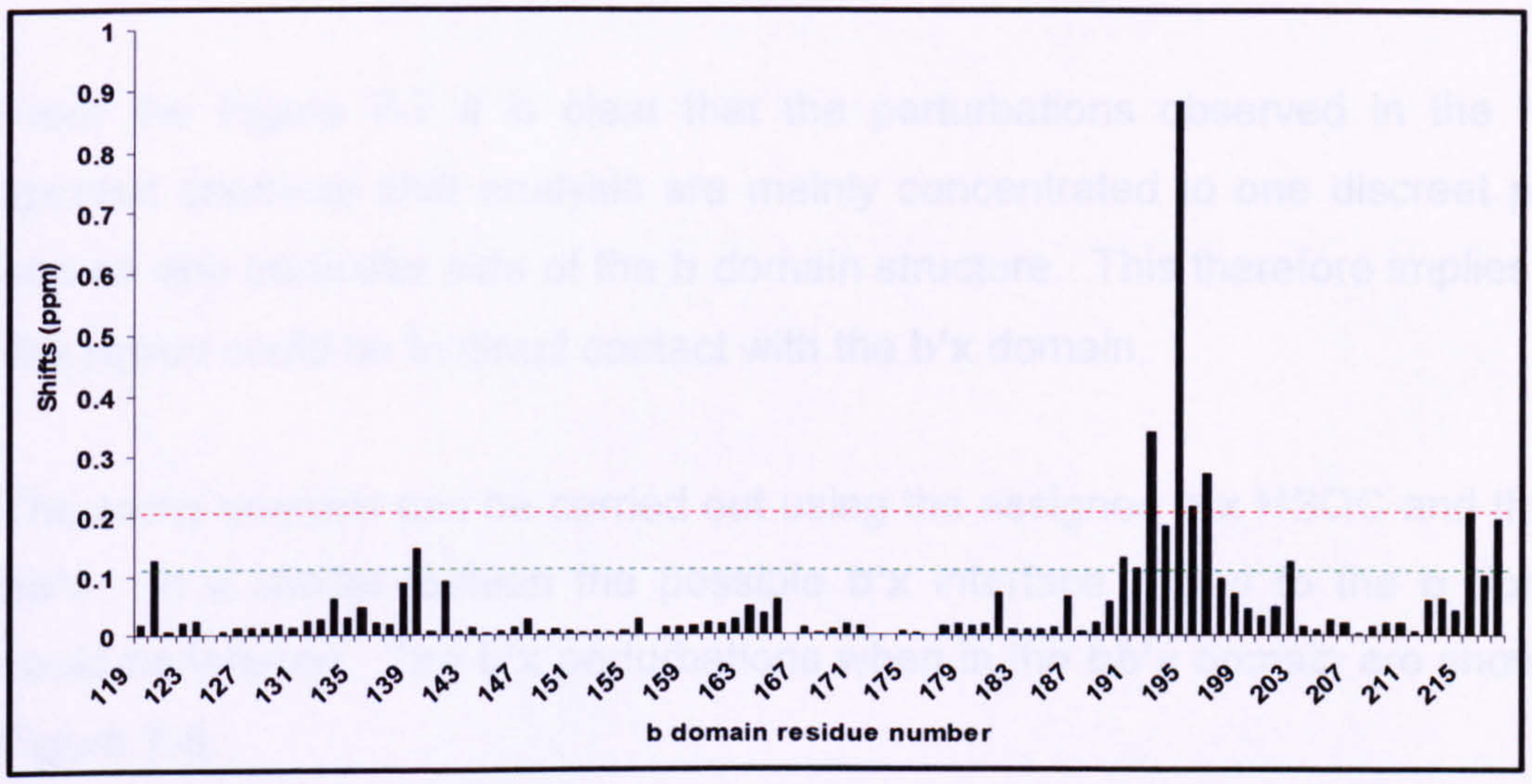

Figure 7-6. Chemical shift perturbation identified in the b domain by comparison of the b domain HSQC and the bb'x HSQC. (---) 1 Standard deviation, (---) 2 Standard deviations.

From the chemical shift analysis shown in Figure 7-6, the residues with greater than 2 standard deviations were Glu194, Phe192, Arg196, Gly195 and Asn215. Those with greater than 1 standard deviation were 217Leu, Asp193, Ala139, Lys190, Glu202, Ala120, Lys191 and Gln216. From the graph in Figure 7-6, it is clear that the perturbation arise in distinct areas; the clearest area is that around residue Asn197. The ends of the domain would inherently be more flexible and reveal high perturbations; hence the area of interest is that around residue 139. These perturbations reveal far more information when plotted onto the **b** domain structure as shown in Figure 7-7.

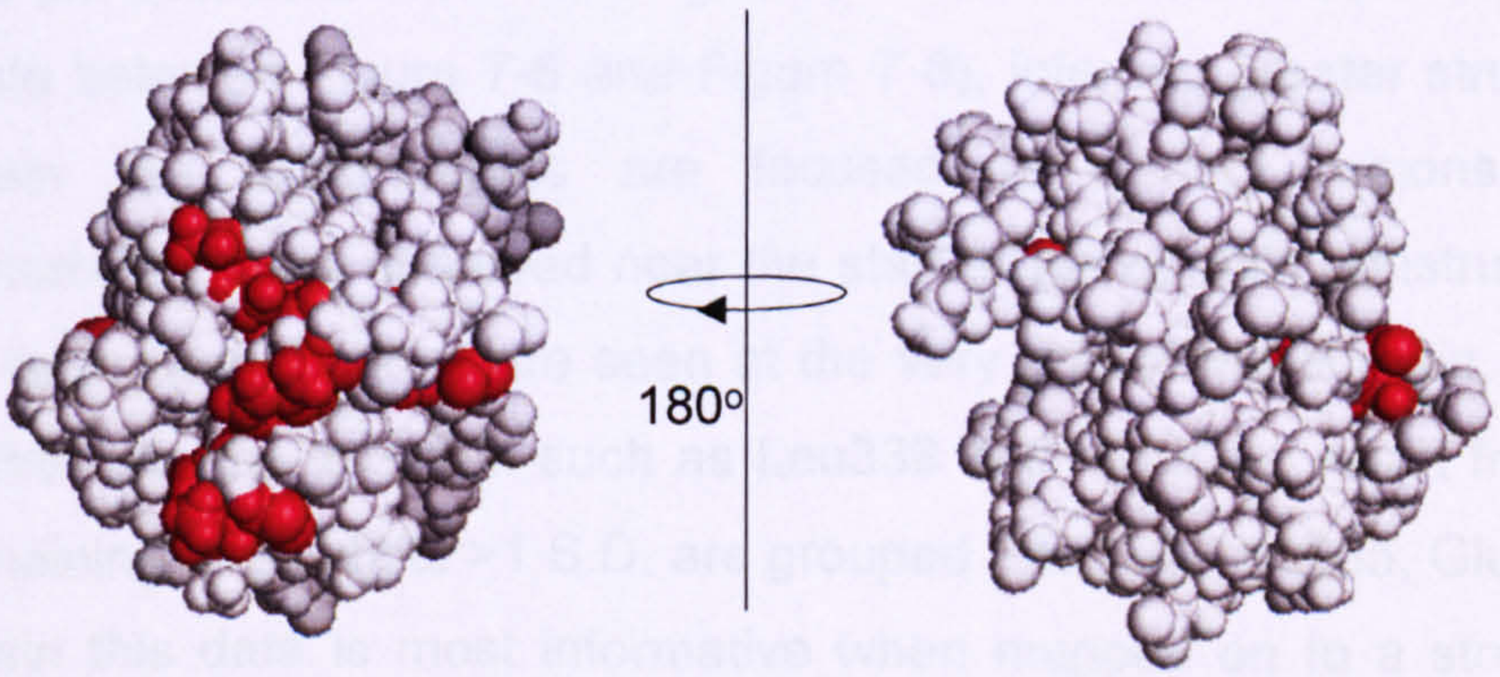

Figure 7-7. Two viewpoints of the space-fill b domain structure. Perturbations >1 standard deviation highlighted in red.

From the Figure 7-7 it is clear that the perturbations observed in the NMR spectral chemical shift analysis are mainly concentrated to one discreet patch and on one particular side of the **b** domain structure. This therefore implies that this region could be in direct contact with the **b'x** domain.

The same analysis can be carried out using the assigned **b'x** HSQC and that of **bb'x**. In a similar fashion the possible **b'x** interface region to the **b** domain could be inferred. The **b'x** perturbations when in the **bb'x** domain are shown in Figure 7-8.

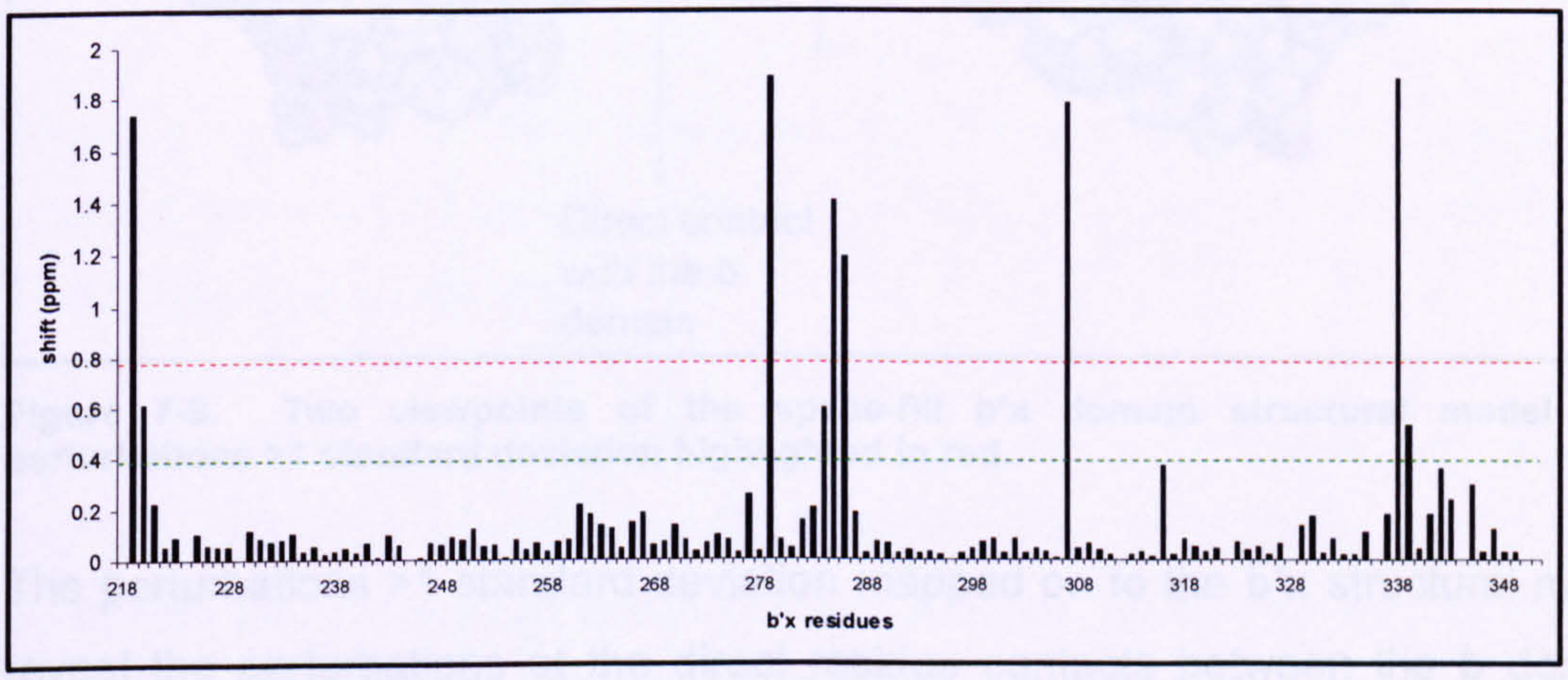

Figure 7-8. Chemical shift perturbation identified in the b'x domain by comparison of the b'x domain HSQC and the bb'x HSQC. (---) 1 Standard deviation, (---) 2 Standard deviations.

The perturbations were much greater in the **b'x** domain (note the difference in scale between Figure 7-6 and Figure 7-8), inferring greater structural flexibility. Again the perturbations are focused to distinct regions, as expected perturbations are observed near the start of the domain construct. Interestingly no large perturbations are seen at the very end of the domain, but seen a little further into the **x**-region such as Leu338 and Met339. Apart from Met307, the remaining large shifts >1 S.D. are grouped (Thr279, Leu285, Glu286 and Ile284). Again this data is most informative when mapped on to a structure, since no human PDI **b'x** structure has been determined a model was produced using the Geno3D (available at http://geno3d-pbil.ibcp.fr) (Combet *et al.*, 2002) which

used the Erp57 **bb'** as a template structure (see section 2.6.1. for more details). This **b'x** model was used to map the perturbations as shown in Figure 7-9.

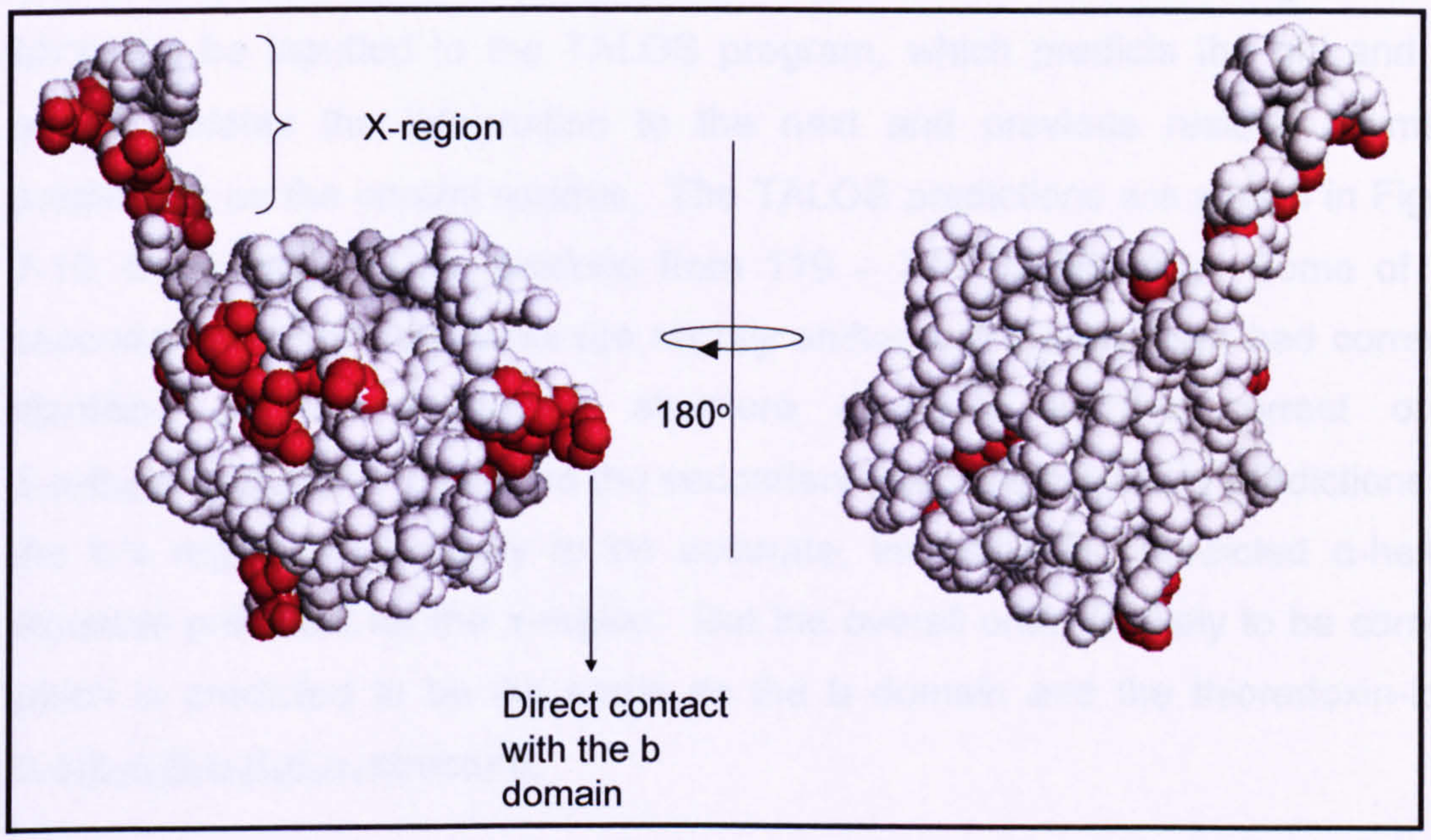

Figure 7-9. Two viewpoints of the space-fill b'x domain structural model with perturbations >1 standard deviation highlighted in red.

The perturbations >1 standard deviation mapped on to the **b'x** structural model reveal the perturbations at the direct residue contacts between the **b** domain and **b'x** domain, highlighted in Figure 7-9. Expected perturbations on the flexible and unbound **x**-region are also highlighted. Importantly residues Thr279, Leu285, Glu286 and Ile284 are grouped to a distinct region and this is expected to be the interface between the domains. Residue Met307 is also perturbed and is also situated on the surface, highlighted in red, at the bottom of the left-hand image in Figure 7-9. At this stage it was unclear if this an internal structural change induced by the addition of the **b** domain or forms part of the interface between the domains.

7.3.3. Secondary structure prediction: TALOS

The Cα, Cβ, HN and Hα chemical shift data for each residue using the 40°C **bb'x** can be inputted to the TALOS program, which predicts the phi and psi angles, relates the information to the next and previous residue to make predictions on the central residue. The TALOS predictions are shown in Figure 7-10, firstly focusing on residues from 119 – 217 (**b** domain). Some of the secondary structure elements are slightly shifted but the program had correctly identified all the secondary structure elements in the correct order $\beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4\beta_4\alpha_4$. Therefore the secondary structure boundary predictions for the **b'x** region are unlikely to be accurate, including the predicted α-helical structure predicted for the **x**-region. But the overall order is likely to be correct, which is predicted to be the same as the **b** domain and the thioredoxin-fold, $\beta_1\alpha_1\beta_2\alpha_2\beta_3\alpha_3\beta_4\beta_4\alpha_4$ structure.

The low accuracy for the TALOS program results from the fact that 40 residues were inputted lacking the Cα and Cβ chemical shifts; 10 residues were prolines for which the Cα and Cβ could not be recorded. The missing Cα and Cβ chemical shifts resulted from poor spectral resolution where no shifts could be recorded.

Figure 7-10. Diagram showing the numbered bb'x sequence, with the TALOS predictions and actual b domain secondary structure. B-strands in red, A-helices in yellow, ends of domains highlighted in grey. X-region helical secondary structure denoted by XA1.

7.4. Hydrogen/Deuterium Exchange

This section describes results collected using hydrogen/deuterium (H/D) exchange experiments on the **b** and **bb'x** domain constructs. Samples remained in high deuterium conditions for 21 hours and spectra were collected at various time points from 5 minutes to 21 hours. The general approach to analysis of this kind of data involves fitting a single exponential decay to residue peaks over the time period. Unfortunately the experiment which was run at 25°C did not give good data and rates were not informative, although a basic analysis is possible by identifying which residues are slow and fast exchanging. The problem of poor data could be resolved by repeating the experiments at 40°C, which should improve spectral resolution.

7.4.1. b domain H/D exchange

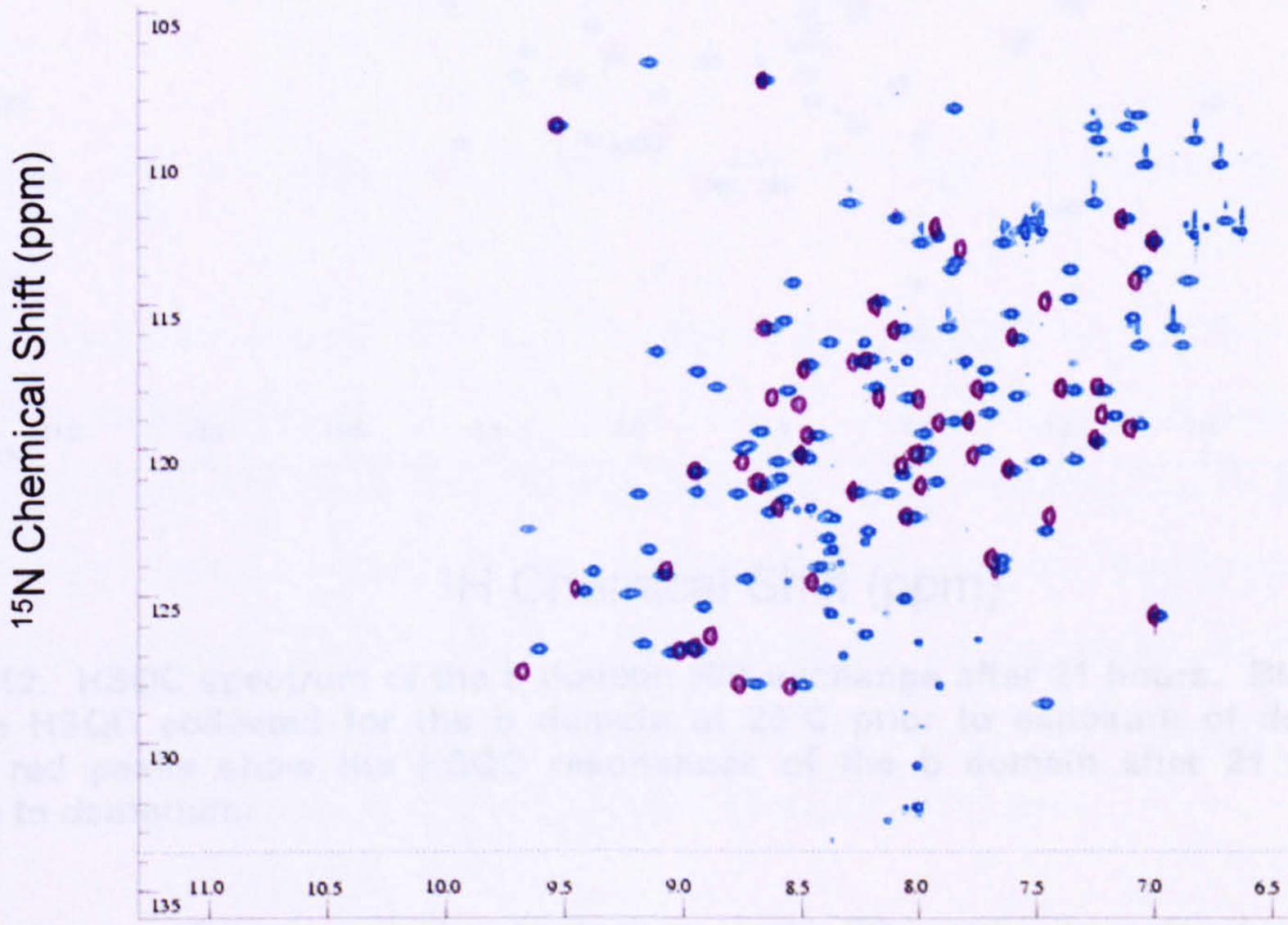

Figure 7-11. HSQC spectrum of the b domain H/D exchange after 5 minutes. Blue peaks show the HSQC collected for the b domain at 25°C prior to deuterium exposure. Overlaid purple peaks show the HSQC resonances of the b domain after 5 minutes of exposure to deuterium.

As highlighted in Figure 7-11, only 52 peaks (44%) remain after 5 minutes of being dissolved in D_2O (listed in Appendix 4). The fast-exchanging 56% of residues are likely to be the solvent exposed residues and regions of structural flexibility. Also noticeable from the spectra is that some of the peaks have shifted slightly, which is likely to result from slight structural changes possibly resulting the freeze-drying process or from deuterium exchange.

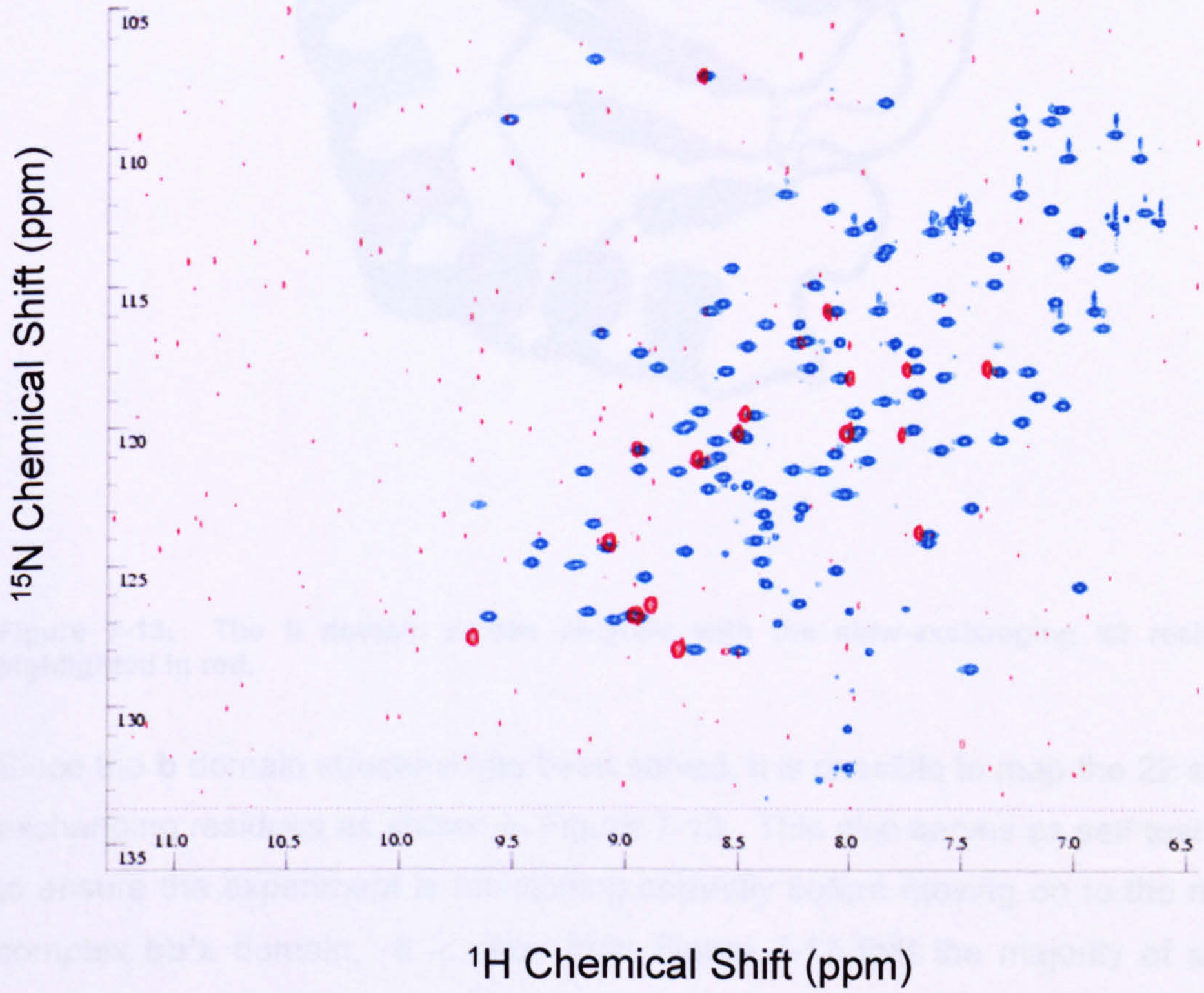

Figure 7-12. HSQC spectrum of the b domain H/D exchange after 21 hours. Blue peaks show the HSQC collected for the b domain at 25°C prior to exposure of deuterium. Overlaid red peaks show the HSQC resonances of the b domain after 21 hours of exposure to deuterium.

After 21 hours of exposure to deuterium, only 22 peaks remained (listed in Appendix 4). These residues are likely to be the most buried residues in the structure or residues situated in stable secondary structure elements and so the least solvent-exposed regions.

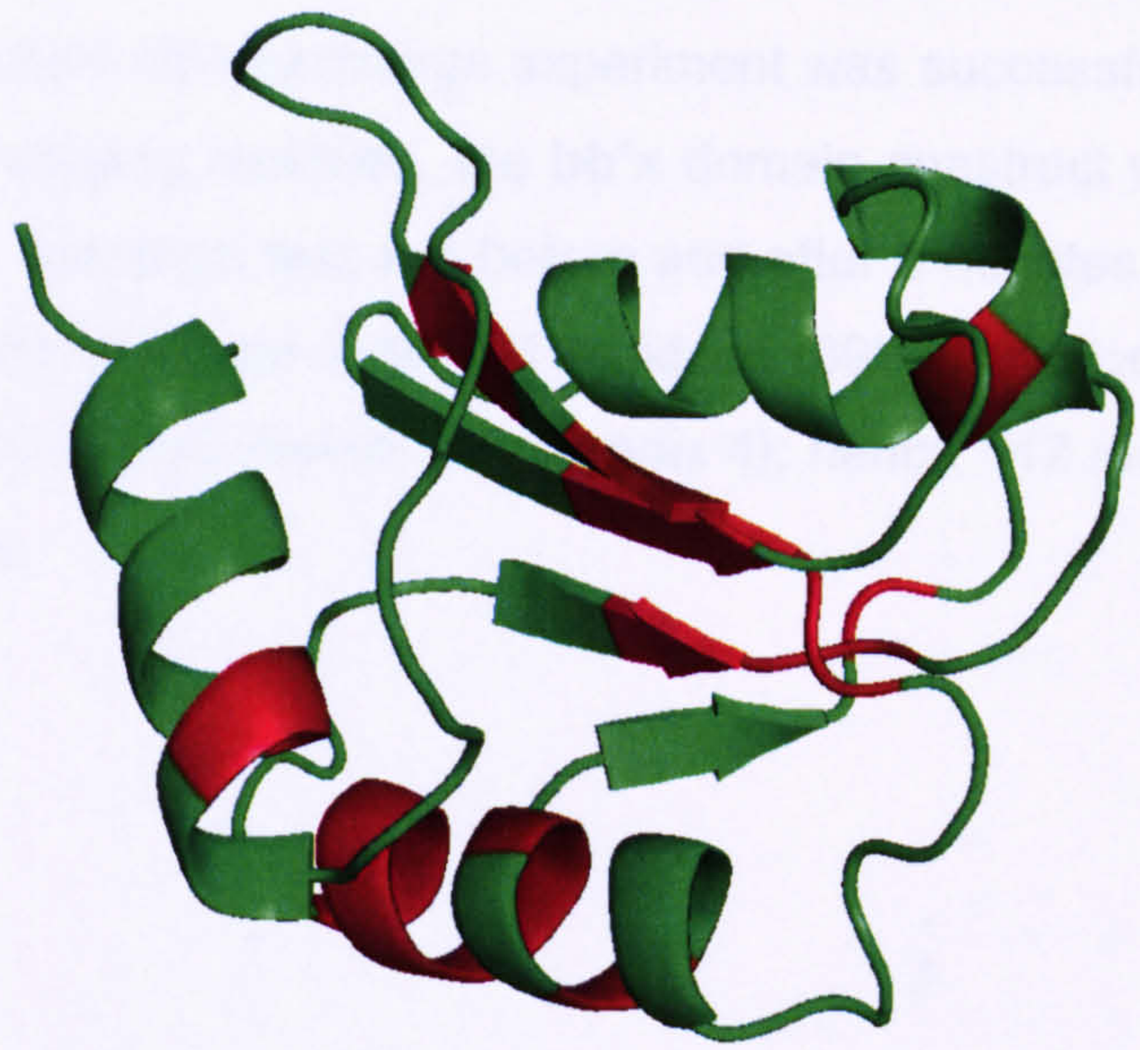

Figure 7-13. The b domain ribbon diagram with the slow-exchanging 22 residues highlighted in red.

Since the **b** domain structure has been solved, it is possible to map the 22 slow-exchanging residues as shown in Figure 7-13. This also serves as self test tool to ensure the experiment is functioning correctly before moving on to the more complex **bb'x** domain. It is clear from Figure 7-13 that the majority of slow-exchanging residues are located at the centre of the molecule, in the least solvent exposed regions. Also highlighted are a number of other residues which as predicted, are located amidst secondary structure elements.

7.4.2. bb'x H/D exchange

Since the **b** domain H/D exchange experiment was successful at identifying the core, slow-exchanging residues, the **bb'x** domain construct was also analysed. Again a HSQC spectrum was run before and after 5 minutes of being dissolved in D_2O, as shown in Figure 7-14. 91 residues (39%) remained after the short 5 minute exposure to D_2O (listed in Appendix 4); hence 142 residues (61%) were fast-exchanging.

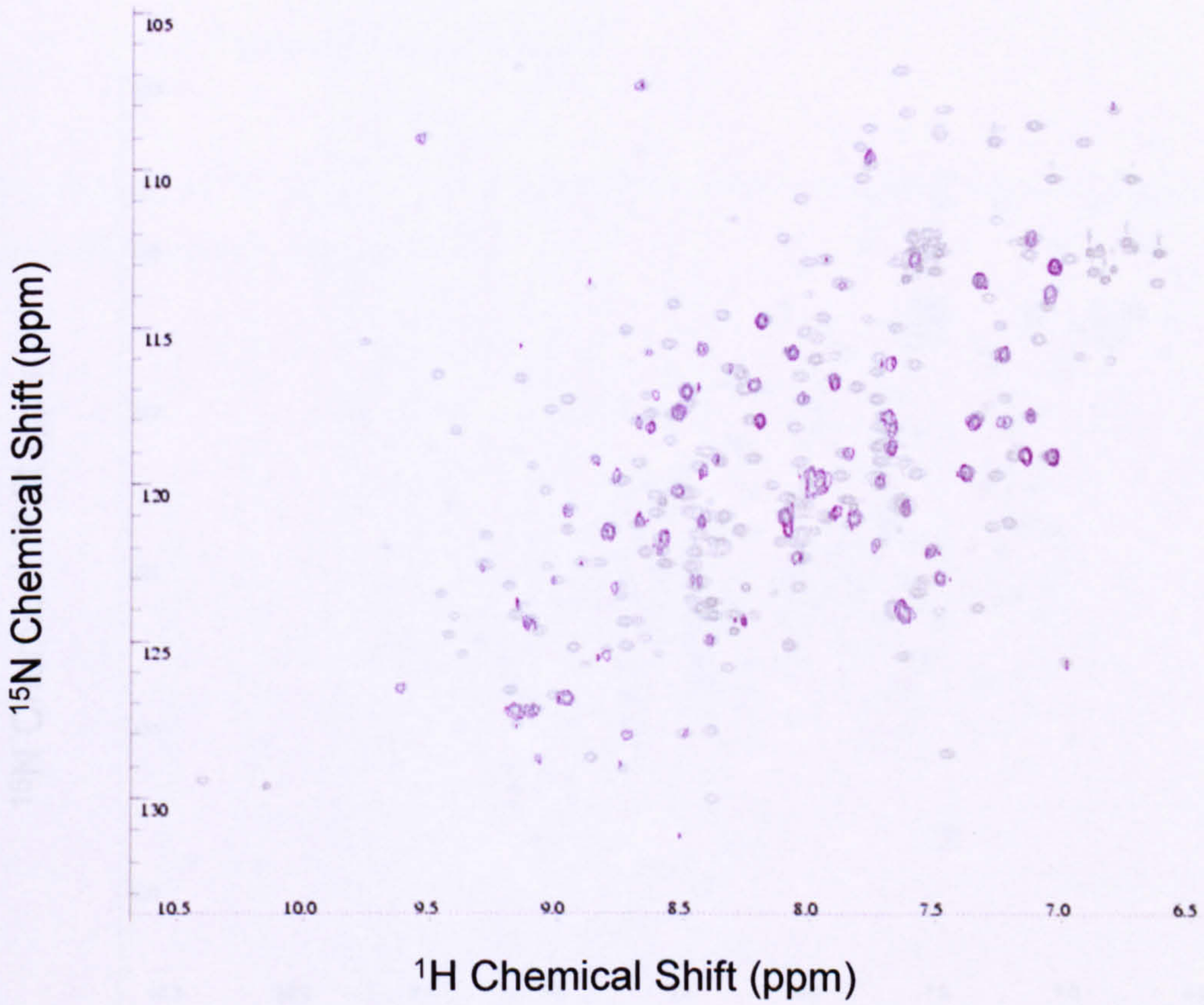

Figure 7-14. HSQC spectrum of the bb'x domain H/D exchange after 5 minutes. Grey peaks show the HSQC collected for the b domain at 25°C prior to deuterium exposure. Overlaid purple peaks show the HSQC resonances of the bb'x domain after 5 minutes of exposure to deuterium.

A HSQC was then re-run after a delay of 21 hours to identify the slow-exchanging residues which are likely to lie at the centre of the protein molecule

and the least solvent exposed regions. As highlighted in Figure 7-15, very few residues remained unexchanged; totalling 34 residues (listed in Appendix 4). Upon comparison of the slow-exchanging **b** domain residues in both H/D exchange experiments of **b** and **bb'x** domains, the same residues were identified. This also validates the confidence in the experimental data. Also of the 34 slow-exchanging residues in the **bb'x** experiment, only 12 residues arise in the **b'x** region of the construct and none in the **x**-region. This suggests the **b'x** and particularly the **x**-region is more exposed as a result of structural fluctuation, which supports the HSQC data described in an earlier chapter suggesting that **b'x** region is more conformationally flexible.

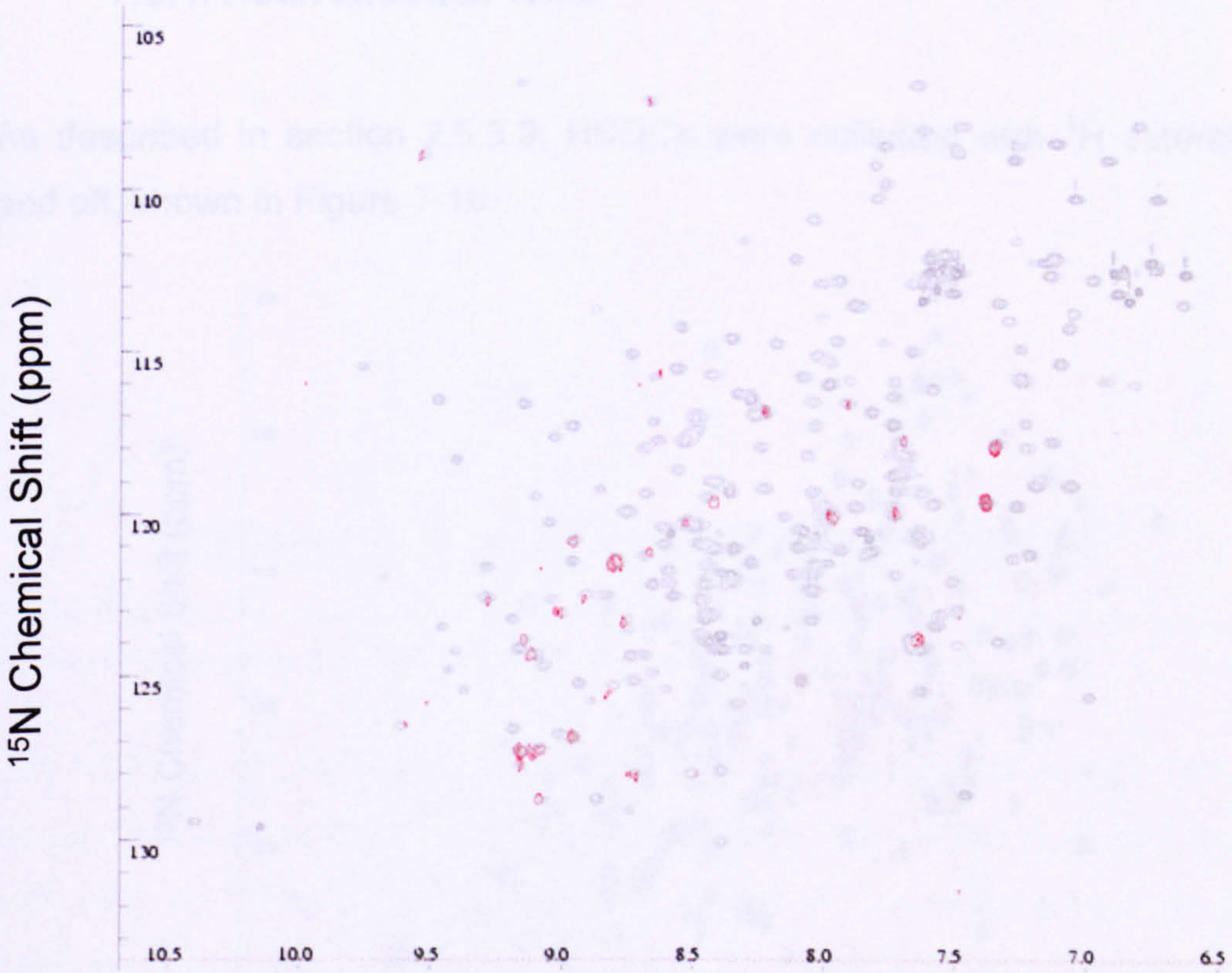

Figure 7-15. HSQC spectrum of the bb'x domain H/D exchange after 21 hours. Grey peaks show the HSQC collected for the b domain at 25°C prior to deuterium exposure. Overlaid red peaks show the HSQC resonances of the b domain after 21 hours of exposure to deuterium.

7.5. Dynamic Analysis using Relaxation Experiments

This section describes more detailed studies involving the collection and analysis of NMR relaxation dynamic (T_1 and T_2) and Het NOE (heteronuclear nuclear overhauser effect) data which not only reveals dynamic information about the whole construct but also structural information on individual domains and residues.

7.5.1. Heteronuclear-NOE

As described in section 2.5.3.9, HSQCs were collected with ^1H saturation on and off; shown in Figure 7-16.

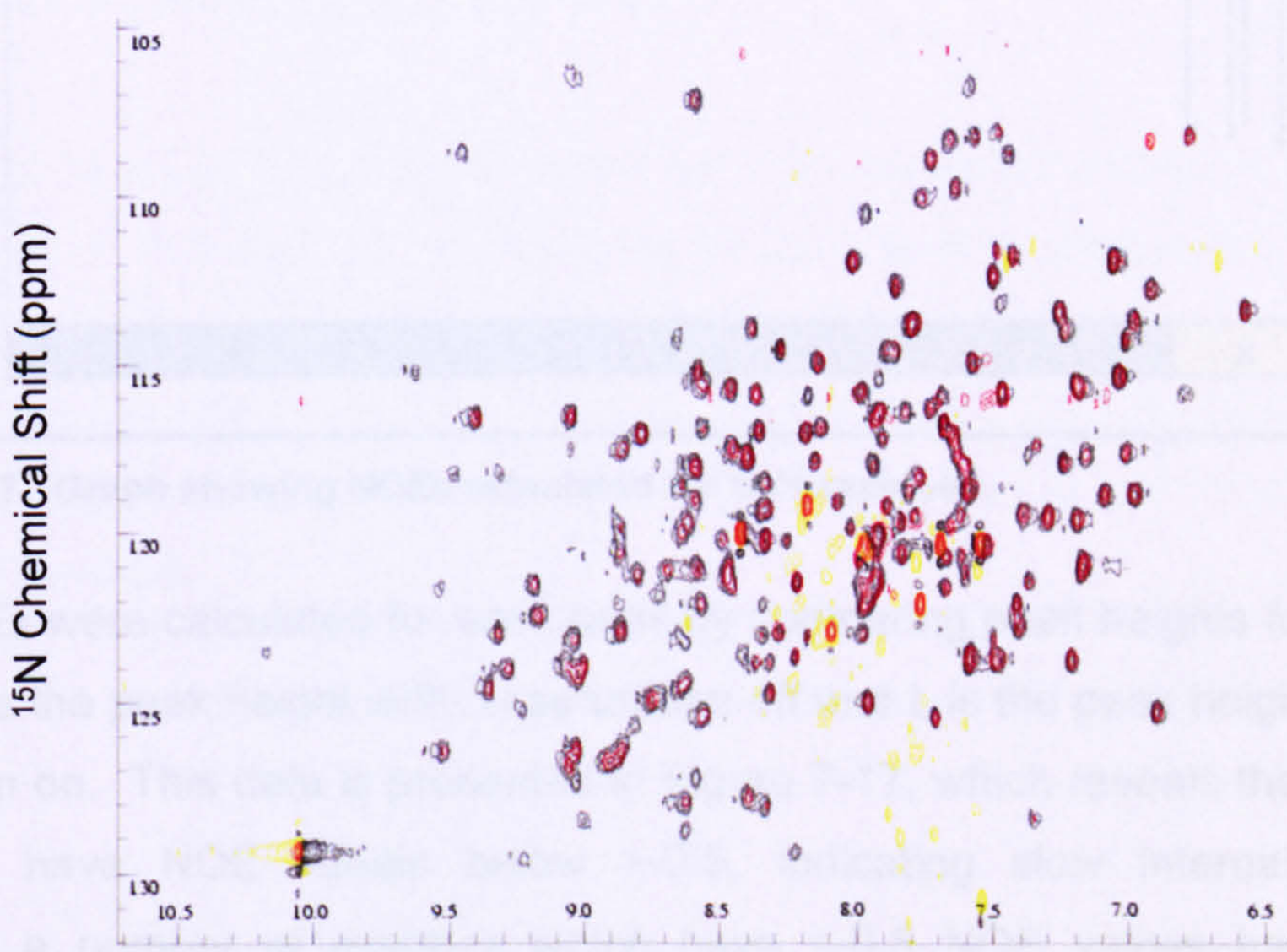

Figure 7-16. 2D-HSQC collected at 40°C of ^{15}N-bb'x with ^{1}H saturation on and off. ^{1}H saturation off is shown in red peaks, ^{1}H saturation on is shown in red peaks for positive peaks and yellow for negative peaks.

As shown in Figure 7-16, with ^{1}H saturation on, many negative are also shown, while the off peaks are weak and broadened, reflecting a slightly poor quality of data collected. Unfortunately due to time constraints, a repeat of the experiment at a lower temperature, which could improve data quality, could not be carried out. But it was deemed the data quality was sufficient to carry out meaningful analysis.

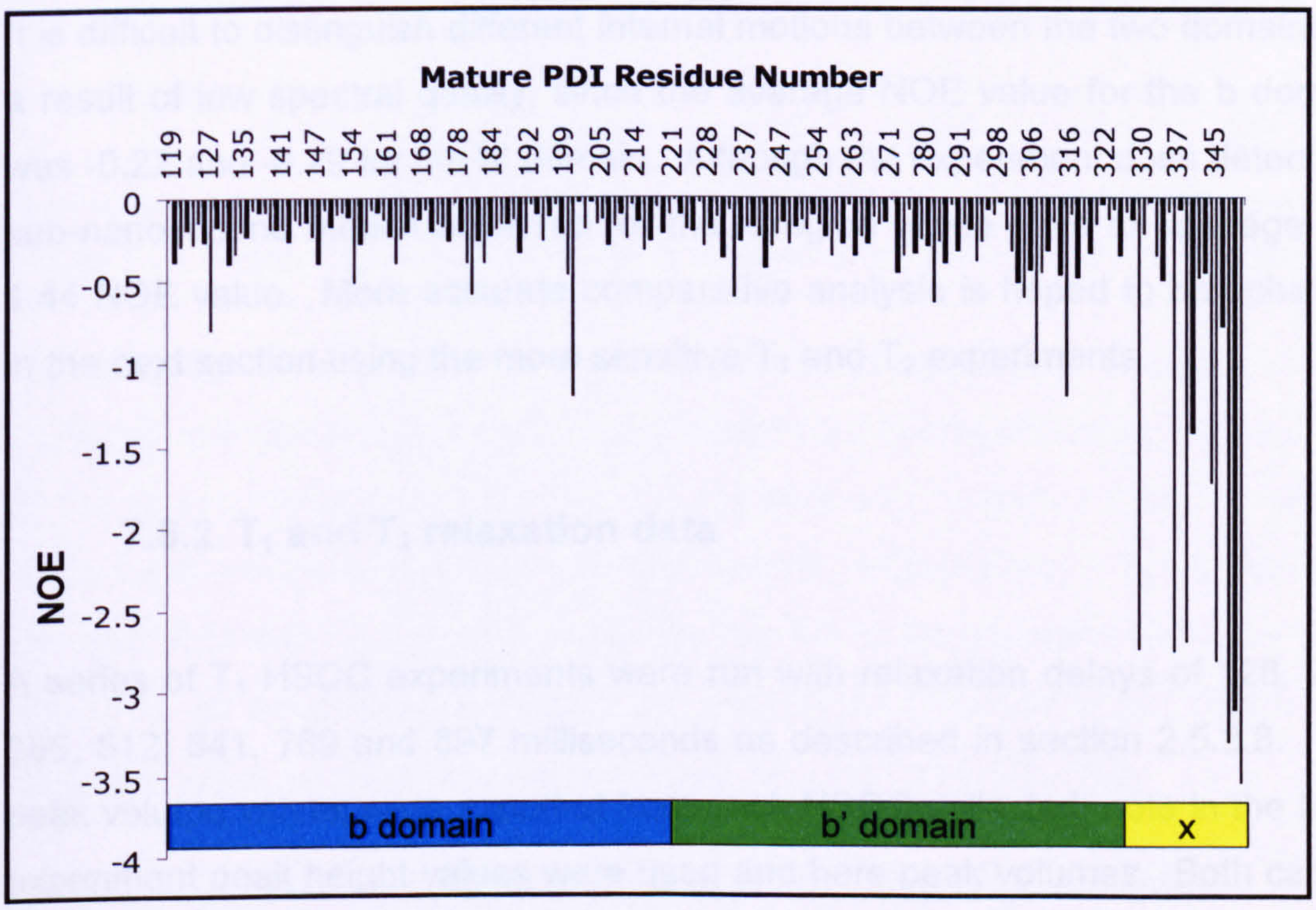

Figure 7-17. Graph showing NOEs calculated for bb'x residues.

The NOEs were calculated for each peak by comparing peak heights for $(I-I_o)/I_o$, where I is the peak height with ^{1}H saturation off and I_o is the peak height with ^{1}H saturation on. This data is presented in Figure 7-17, which reveals the majority residues have NOE values below >-0.5, indicating slow internal motion. However a number of residues which have <-0.5 NOE values have faster internal motion and were found to lie in loop regions when mapped on to the **b** domain structure and therefore it was assumed that residues with <-0.5 NOE values in the **b'** domain are also likely to be situated in loop regions. Also very noticeable was the very low NOE values recorded from the **x**-region, indicating

a very high degree of rapid motion, which may be a feature of the region or possibly resulting from termini flexibility, which is unlikely as the low values extend throughout the final 20 residues. This motion occurs in a longer region than expected for only termini flexibility which normally only extends for a few residues into the protein at terminal ends; an example of this is highlighted by Mackay and colleagues (MacKay *et al.*, 1996).

It is difficult to distinguish different internal motions between the two domains as a result of low spectral quality; since the average NOE value for the **b** domain was -0.22 and -0.29 for the **b'** domain. Although the experiment does detect the sub-nanosecond motion observed for the **x**-region which gave an average of -1.44 NOE value. More accurate comparative analysis is hoped to be achieved in the next section using the more sensitive T_1 and T_2 experiments.

7.5.2. T_1 and T_2 relaxation data

A series of T_1 HSQC experiments were run with relaxation delays of 128, 256, 385, 512, 641, 769 and 897 milliseconds as described in section 2.5.3.8. The peak volume values were exported from each HSQC collected; note in the NOE experiment peak height values were used and here peak volumes. Both can be used to analyse both sets of experiments, with the choice dependent on the quality of the data.

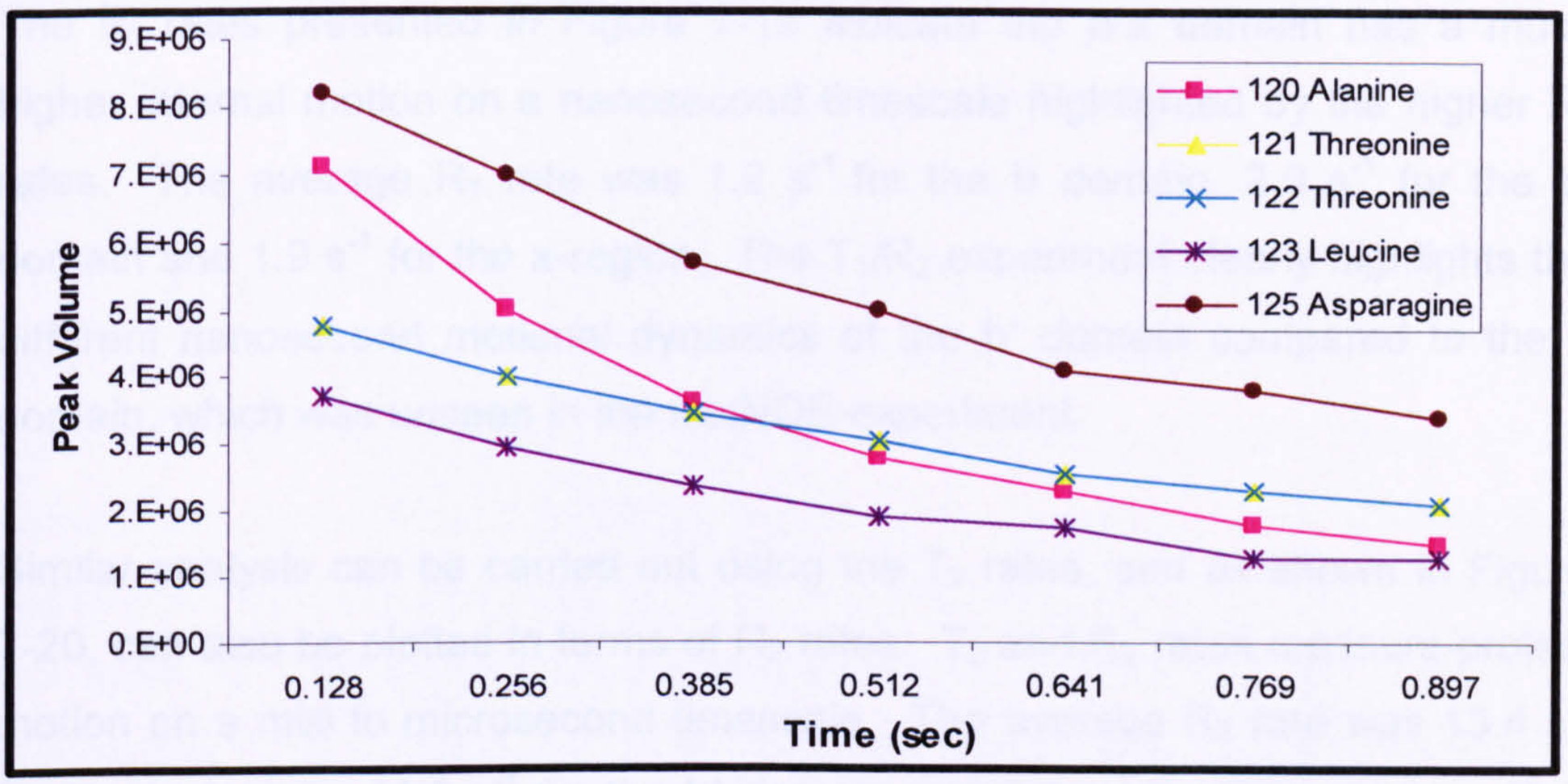

Figure 7-18. Typical T_1 decay for 5 residues using peak volume against T_1 time delays.

The peak volumes for each residue can be plotted against the T_1 delay and the T_1 decay rate calculated for each residue from the three-parameter nonlinear fits of the experimental data to $I(\tau)=I_\infty - [I_\infty - I_o] \exp(-\tau/T_1)$, where , I_o and I_∞ are the cross-peak volumes at zero- and infinite-time, respectively. T_2 values were likewise obtained from the three-parameters fits to the equation $I(\tau)=I_\infty + [I_\infty + I_o] \exp(-\tau/T_2)$ as described by Mackay and colleagues (MacKay *et al.*, 1996). The T_1 and T_2 rates are regularly expressed as $1/T_1$ (R_1) and $1/T_2$ (R_2).

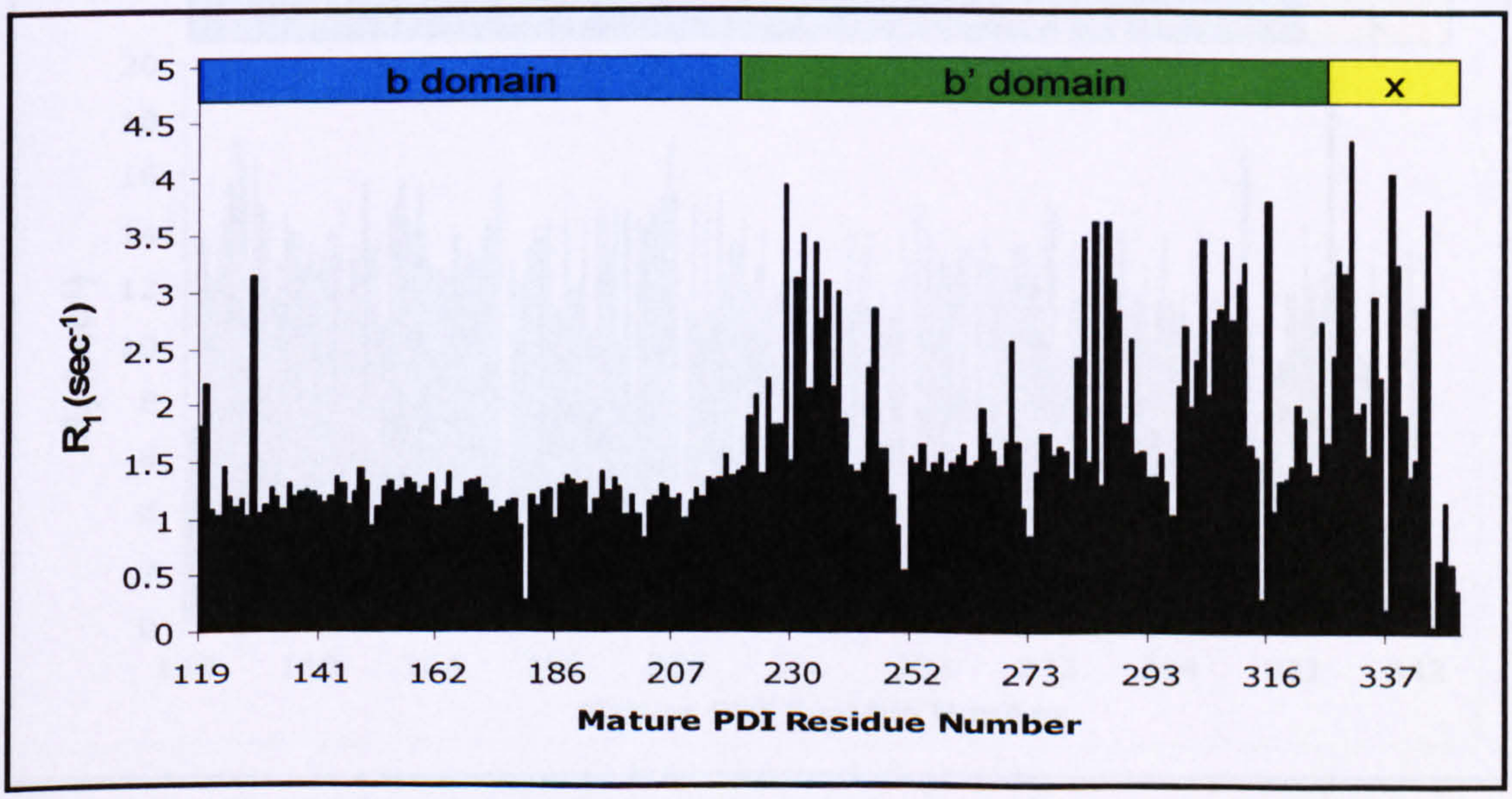

Figure 7-19. R_1 rates for bb'x residues.

The R_1 rates presented in Figure 7-19 indicate the **b'x** domain has a much higher internal motion on a nanosecond timescale highlighted by the higher R_1 rates. The average R_1 rate was 1.2 s^{-1} for the **b** domain, 2.0 s^{-1} for the **b'** domain and 1.9 s^{-1} for the **x**-region. The T_1/R_2 experiment clearly highlights the different nanosecond motional dynamics of the **b'** domain compared to the **b** domain, which was unseen in the HetNOE experiment.

Similar analysis can be carried out using the T_2 rates, and as shown in Figure 7-20, can also be plotted in terms of R_2 rates. T_2 and R_2 rates measure protein motion on a milli to microsecond timescale. The average R_2 rate was 13.4 s^{-1} for the **b** domain, 11.5 s^{-1} for the **b'** domain and 5.8 s^{-1} for the **x**-region. This does not highlight any distinct differences between the **b** and **b'** domain at these timescales; but supports the results seen in the R_1 analysis where motion is observed, and must be on the faster nano-picosecond timescale. The very low **x**-region R_2 average can be explained since only the very terminal residues indicate a different timescale of motion which is likely to be attributable to termini flexibility and so give the very low R_2 rates. But not all the **x**-region residues have very low R_2 values hence the low R_2 average is biased by the termini flexibility.

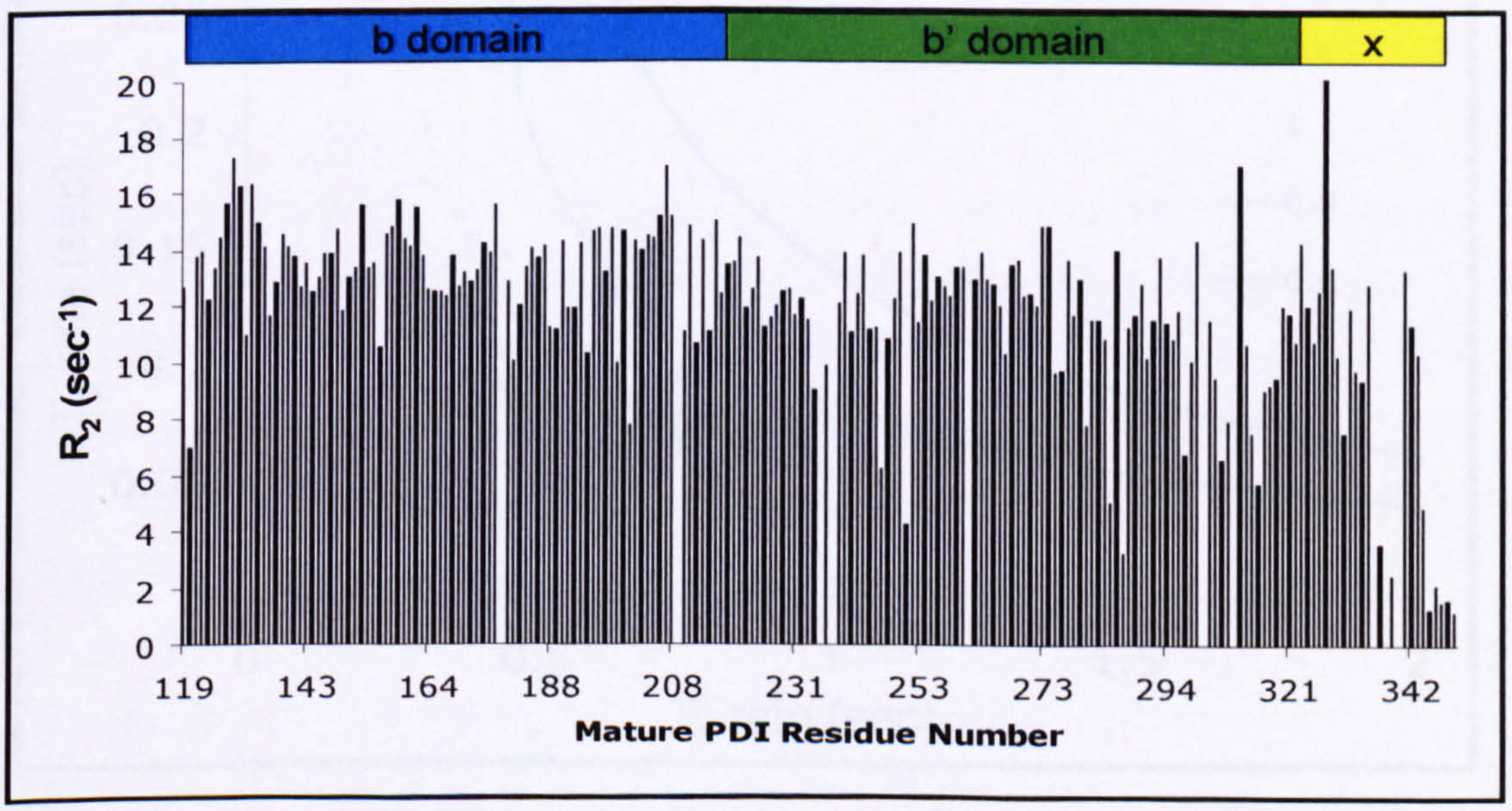

Figure 7-20. R_2 rates for bb'x residues.

7.5.3. T_1 and T_2 rates to determine isotropy or anisotropy

Using the model-free formalism (Lipari and Szabo, 1981; Lipari and Szabo, 1982) contours can be calculated with spectral density functions to represent fixed values for the order parameter S^2 with varying correlation times τ_m. These theoretical contours are shown in Figure 7-21 and are calculated based on an internal correlation time τ_e of 50 picoseconds. If T_1 and T_2 rates are recorded above the theoretical S^2 value of 1 then the protein is predicted to be spherical and isotropic, whereas any rates recorded below S^2 value of 1 are likely to be anisotropic. From Figure 7-21 it is clear that the majority of **b'x** residues fall below the S^2=1 contour and so indicating that this domain is anisotropic. Whilst the majority of the **b** domain residues lie above the S^2=1 contour and so is likely to isotropic. As a result of the anisotropic nature of the **b'x** domain, use of the Lipari and Szabo Model-free approach to interpret protein dynamics and determine dynamic parameters was not possible.

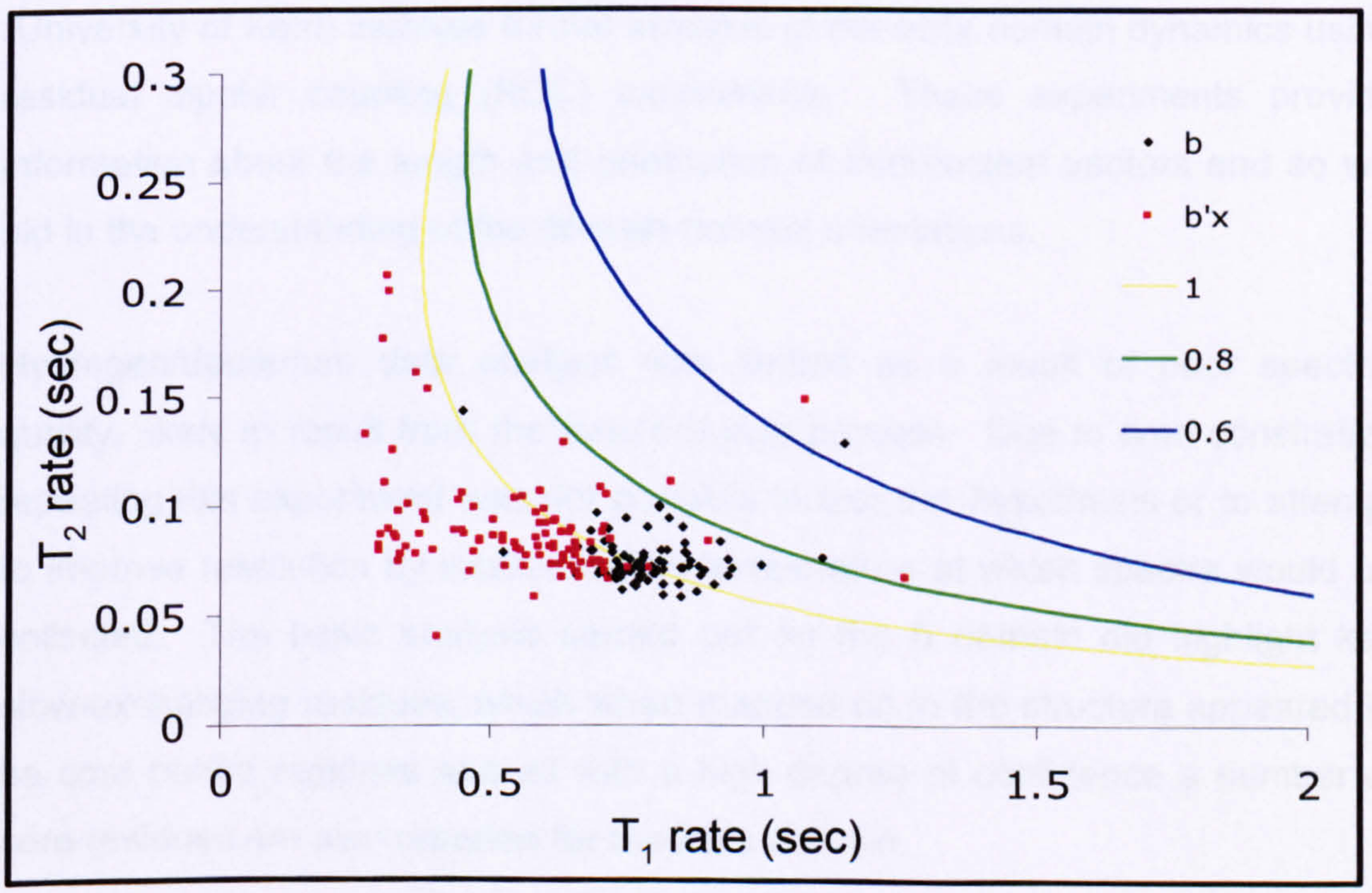

Figure 7-21. Plot of experimental T_1 against T_2 rates. The yellow, green and blue contour lines represent theoretical 1, 0.8 and 0.6 S^2 rates, respectively.

7.6. Discussion

The chemical shift analysis not only allowed the transfer of assignment to the HSQC spectrum collected at 25°C, at which temperature backbone assignment was impossible, the assignment allowed more probing dynamic and detailed molecular NMR analysis. Since 25°C is the standard NMR experiment running temperature the ability to assign a HSQC spectrum at this temperature was vital. Analysis of chemical shifts was very informative, firstly by analysing chemical shifts over the temperature changes of 40°C to 25°C revealed a number interesting residues likely to be structurally significant in allowing protein flexibility. Secondly, by mapping chemical shifts between spectra from different domain constructs. This led to very convincing evidence of a contact site on the **b** domain which interacts with the **b'x** domain. The data also suggested that the interaction of the **b'x** domain with the **b** domain causes long-range internal structural changes. Also, current on-going work carried out by Dr. L. Byrne (University of Kent) involves further analysis of the **bb'x** domain dynamics using residual dipolar coupling (RDC) experiments. These experiments provide information about the length and orientation of internuclear vectors and so will aid in the understanding of the domain-domain orientations.

Hydrogen/deuterium data analysis was limited as a result of poor spectral quality, likely to result from the freeze-drying process. Due to time constraints repeating this experiment was not possible to test this hypothesis or to attempt to improve resolution by increasing the temperature at which spectra would be collected. The basic analysis carried out on the **b** domain did highlight key slow-exchanging residues, which when mapped on to the structure appeared to be core buried residues and so with a high degree of confidence a number of core residues are also reported for the **bb'x** domain.

The T_1 and T_2 relaxation experiments highlighted a distinct difference between backbone motions of the **b** and **b'x** domains. Unfortunately, since the **b'x** domain was identified to be anisotropic, quantification of backbone motion using the model-free formalism was not possible. Understanding the complex nature of backbone motion for the **b'x** domain will require further NMR dynamic analysis, possibly using the 'extended model-free' approach (Clore *et al.*, 1990), designed for intradomain dynamics comprising of fast and slow motions. Alternatively, more recently a method of modelling interdomain dynamics was proposed by describing the dynamics as an interconversion between two distinct states (Ryabov and Fushman, 2006).

Chapter 8. Discussion

8.1. Introduction

In previous work on PDI, scientific attention had been focused primarily on the **a** and **a'** domains which were shown to contain the active site C-X-X-C consensus sequences; involved in disulphide bond catalysis. Attention later turned to the **b** and **b'** domain since it was shown that these were required for the complex isomerisation reactions (Darby *et al.*, 1998). It was also found that the **b'a'c** construct was the minimum PDI domain combination that allowed isomerisation and the addition of the **b** domain to this construct increased the activity further. This interest led to the structural determination of the **b** domain in 1999 by the Kemmink group, but attempts to crystallise or obtain high resolution NMR data on the **b'** domain were unsuccessful. Hence it was not until 2006, two years into this research project, that the first **b'** domain structure had been determined; highlighting the intransigent nature of the **b'** domain in structural analysis. At that time, the structures of the three other human PDI domains, **a**, **b** and **a'** had been determined (Kemmink *et al.*, 1996; Kemmink *et al.*, 1999; Tochio *et al.*, to be published) as had other **a** and **a'** domains from a host of different species. The importance of the **b'** domain was established when experimental data suggested a principal substrate binding site lay in this domain and proved the functional significance of the **b'** domain in PDI function (Klappa *et al.*, 1998).

Hence this study set out initially with an ambitious goal of full structural determination of the **b'** domain which was later limited to obtaining backbone assignments for a **b'** domain containing construct. The procedure to make assignments was made difficult by poor NMR data resolution and an elaborate method using a range of NMR experimental data was required to achieve the goal. The assignments were used to analyse structure and backbone dynamics.

8.2. Expression and Characterisation of PDI Proteins

A detailed protocol, involving sample fractionation, was established in this research to allow the preparation of homogenous protein samples. This provided the basis to obtain quality NMR data, improving upon prior preparations of samples. Fractionation was based on the observation of two species during the purification steps, which was initially speculated and later characterised to result from a monomer and dimer species. The dimer species was shown to be a non-covalent dimer, but upon NMR analysis this pool was later defined to also contain the open form of the **bb'x** molecule and the monomer defined as a closed form; whereby the open and close refers to the positioning of the **x**-region. Generally the 'monomer' sample was shown by native-PAGE to be a homogenous sample. The 'dimer' sample appeared to be less homogenous; this was also shown in NMR experiments whereby spectra appeared to show a poorly resolved monomer species. This sample was defined as the open form of the **bb'x** species, evidence of this open and close form was further analysed in the **b'x** domain (Nguyen *et al.*, (Manuscript in preparation)), see Appendix 1. Unfortunately a sample of the dimer pool was not analysed by AUC during this study as attention was focused on the monomer species; this would be an important experiment for future work to quantify the amount of non-covalent dimer produced. The dimer pool also requires further characterisation to see if the dimer and open forms of **bb'x** domain can be separated.

Expression particularly of the **bb'x** domain was optimised to reduce the amount of ^{13}C-glucose used in the minimal expression medium, whilst maintaining the high levels of protein yields required for NMR analysis. It was also shown by mass spectrometry analysis that high levels of isotopic labelling could be achieved using ^{15}N and ^{13}C labelled reagents in minimal growth medium. The

most important protocol established in this research defines a procedure to produce ^{2}H, ^{15}N, ^{13}C-labelled proteins, we foresee no reason why this protocol could not be used for other PDI domain combinations and it is likely that any future work on large PDI domain combinations will require triple-labelling to overcome the low spectral resolution from analysis of large protein constructs.

8.3. NMR Approach to Obtain b' Backbone Assignments

Chapter 4 describes prior NMR analysis carried out by Dr. R. Williamson and initial NMR analysis using samples produced early on in this study. The spectral analysis carried out in Chapter 4 was used to devise an appropriate approach to enable backbone assignment of the **b'** domain. Since the **b'** domain was the main domain of focus, collecting NMR spectra on this domain would be the easiest and most obvious starting point. Unfortunately the **b'** domain gave very poor NMR data, as measured using average peak width analysis, described in section 5.2; believed to result from conformational exchange which is likely to be an intrinsic feature linked to function in terms of substrate binding.

Improvement to spectra was observed with the addition of domains, as seen with the addition of the **b** domain in the **bb'** domain construct. A more marked improvement was seen with addition of the **x**-region in the **bb'x** domain construct. This was believed to be caused by the **x**-region binding to the **b'** domain, stabilising this domain and so resulting in an improved spectrum. This hypothesis was supported once full assignments were obtained; the **x**-region peaks in the HSQC were much stronger than any other peaks by a substantial amount. This implies the **x**-region is indeed held in stable conformation but no NMR experimental data obtained here defines the site of the binding, which is still speculated to be the **b'** binding site since the peak resolution of this domain were improved. Future work would entail titrating a ligand into a **bb'x** sample

and recording perturbations to HSQC spectrum which would highlight specific residues involved in substrate binding. Presently there is no known or well established substrate molecule; all reported work on substrate binding has been carried out using short hydrophobic peptides with the exception of BPTI. A possible approach could be to synthesise the **x**-region and use this short peptide as substrate molecule; since the NMR experiments reported here shown the **x**-region binding in the **bb'x** PDI domain combination.

It is important to note that here it is suggested that the **x**-region is bound to the **b'** domain and later, paradoxically, the relaxation experiments suggested the **x**-region had flexibility. This can be explained by the fact that the HSQC is influenced more by the stable **x**-region conformation, which results in stronger HSQC peaks; the weaker extra peaks believed to result from the unbound **x**-region were observed but were so weak and poorly resolved as a result of their motion that they were impossible to assign. Conversely the relaxation experiments are influenced more by the unbound **x**-region, the bound form would influence to a certain degree but not enough to suppress rates to a negligible level.

From the observed improvement to the **bb'x** spectrum, it was logical to target the **b'x** domain which would reduce assignment complexity and it was hoped that the **x**-region would again stabilise the **b'** domain to give a well resolved HSQC spectrum. Unfortunately, the HSQC spectrum was not initially radically improved as seen with the **bb'x** spectrum, this may be caused by the fact the **b'x** sample was unfractionated; whereas the **bb'x** was fractionated based on anion exchange chromatography.

A major breakthrough was the protocol designed in this research to fractionate the samples into two pools. As described in section 5.3, the monomer fractionated pool again further improved the HSQC resolution; to such a degree that it was hoped assignments could be made. A key feature observed in the

HSQC spectrum, was the double tryptophan signal resulting from the single tryptophan residue situated in the **x**-region. Although this signal was observed and noted in early NMR spectra collected by Dr. R. Williamson, only in this study using the fractionated samples could a link be made as to the origin of this signal. The left-hand indole side chain peak was clearly linked to the monomer pool fractions, hence indicating and supporting the hypothesis of the **x**-region binding to the **b'** domain, since this fraction resulted in a highly resolved HSQC spectrum; indicative of a high stable molecule. Whereas a stronger right-hand indole side chain peak, gave very poorly resolved HSQC indicating a conformationally flexible molecule, likely to result from an unbound **x**-region. The important outcome resulting from the stringent fractionation was that the HSQC resolution was now sufficiently high enough that backbone assignment could be possible.

The effect and nature of the **x**-region is intriguing; in work carried out by Dr. K. Wallis and Dr. L. Ruddock, this behaviour has been characterised in the **b'x** domain construct and work based on this feature has enabled the structural determination of a **b'x** mutant which has the **x**-region constantly fixed in the bound state with the **b'** domain (Nguyen *et al.*, (Manuscript in preparation)), see Appendix 1. Ongoing work carried out by the Freedman, Ruddock and Williamson groups focuses on defining the importance of the **x**-region features in the full length human PDI molecule.

The most important and relatable structural information was the recent publication of the human **bb'** domain NMR chemical shifts assignments (Denisov *et al.*, 2006) two years into this study and at a time when the chemical shift assignments reported here were almost complete. The chemical assignments are from human PDI residues 135-357, this construct has an extra 5 residues on the N-terminus which are not found in the human PDI sequence. But more importantly, the construct has only the first 7 residues of the widely accepted **x**-region (Freedman *et al.*, 1998) and lacks the final 12 residues of the

x-region. This is significant as the **x**-region appears to have a unique involvement in the domain construct and may affect the biological relevance of this domain construct. Hence dynamic and structural analysis on an untrimmed **x**-region in the **bb'x** domain as described in this study is likely to yield more relevant data.

Assignment of the **bb'x** was not a simple task as first expected, even though the **bb'x** HSQC was highly resolved, the triple resonance experiments lacked the similar high resolution. As discussed in section 6.4, an elaborate method requiring three data sets (**b** domain, assigned by Dr. K. Wallis, **b'x** assignments and **bb'x** TROSY data) was used to enable full backbone assignments of the **bb'x** domain. Spectra were also required to be collected at 40°C; this higher temperature increased the molecular tumbling rate and so resulted in improved spectra quality, assignments of spectra at 40°C which could then be mapped back to the 25°C HSQC. Assignment was a labour intensive procedure, requiring many iterations of assignments in each data set to ensure no mis-assignments were made. This task was carried out by myself and Dr. L. Byrne, whereby assignments were made independently and then cross-checked, to again minimise mis-assignments. The main difficulties were peak overlap and the low resolution of the triple-resonance data, caused primarily by the size of the protein. The **bb'x** domain is 27477 Da, which is not too large for NMR analysis but due to the protein's conformational flexibility resulting in peak broadening and overlap, the protein bordered on the limits of NMR analysis feasibility. NMR protein analysis is plagued by this size limitation and spectral complexity issues; although using a number of NMR experimental data sets is not novel, it is expected that for the larger and more demanding proteins of biological importance such elaborate approaches will be required.

8.4. Protein Conformation and Dynamics

Backbone assignments are the first step in detailed analysis, and if a demanding protein molecule is analysed, as described here, a usually mundane and simple task as backbone assignment can become a complex problem requiring ingenuity and lateral thinking. The product of the immense effort to obtain NMR assignments for **bb'x** and **b'x** was that they pave the way for more detailed structural and dynamic analysis.

At the onset of this study no full length PDI structure was available, but again two years into this study the first full length PDI structure from *Saccharomyces cerevisiae* was published (Tian *et al.*, 2006). But still no full length human PDI structure has been reported to date. NMR structural determination of any full length PDI protein was impossible due to the large size of the PDI molecule (55kDa). Hence the long term strategy devised at the onset of this project by the three groups involved would be to collect NMR assignments for small domain constructs and use these in a similar fashion as used in this study to define interfaces between domains. Strong evidence of the residues involved in the interface between the **b** and **b'x** domain were found. So by using additional NMR data, primarily residual dipolar coupling experiments to define domain orientation, a full length PDI domain map could be produced; overcoming the size limitations to define the full length PDI structure. This remains a target for the groups and is an area of future work to be done.

The basic analysis using the hydrogen/deuterium exchange experiments revealed important core buried residues in the **b** and **bb'x** domains. Unfortunately the data quality was poor and exchange rates were not accurate. A repeat of this experiment is required, run at a higher temperature which could overcome the problems of poor data quality.

The T_1 and T_2 relaxation experiments were very enlightening, highlighting the significantly different motions in the **bb'x** domain. These experiments highlighted the high degree of flexibility and motion of the **x**-region. The relaxation data therefore suggests that the **x**-region is bound, but can open and close; hence has flexibility. The observed improvement to HSQC spectra results from the fact that when the protein molecules are in an open state the signals are very weak as a result of movement, whereas the closed state would give stronger signals and drown out the weaker open state resonances. Unfortunately, since the **b'x** domain in the **bb'x** domain construct was identified to be anisotropic, quantification of the backbone motion using the model-free formalism was not possible. Understanding the complex nature of backbone motion for the **b'x** domain will require further NMR dynamic analysis, possibly using the 'extended model-free' approach (Clore *et al.*, 1990) or modelling the interdomain dynamics as an interconversion between two distinct states (Ryabov and Fushman, 2006). Current ongoing work includes collecting residual dipolar coupling (RDC) experimental data which will further define interdomain orientations. The dynamic analysis carried out to date is in its infancy and is likely to further quantify the differences in motions observed in this study. This analysis relies on the backbone assignments, which was the most significant achievement described in this study. The backbone assignments provide the springboard to allow more highly detailed structural and dynamic analysis, which is hoped to lead to a greater understanding of the **b'** domain substrate binding function and synergic relationship with the **b** domain.

All the dynamic and structural data collected during this project reveals the **b'** and **b** domains have intrinsically different backbone motions. The dynamic analysis of the **b** domain revealed faster motion than the **b'** domain. This is hypothesised to be related the protein function, where the slower motion in the **b'** domain could be related to the substrate binding function. Where the **b'**

domain could be interpreted as breathing and feeling it surrounding space for a substrate molecule. Future experiments involving titration of a substrate could reveal differences in backbone motion upon substrate binding. As discussed in section 1.3.5, there are distinct differences between human and the recently published yeast PDI structure. The results here and proposed future analysis allows a more versatile and definitive method to analyse the affect of substrate binding on backbone motion which is expected to greatly differ from the Yeast PDI.

Furthermore since a full length PDI structure has eluded research groups, defining interfaces between domains as described in this research used in conjunction with the proposed future RDC analysis could allow us to build a model of the full length PDI protein. The interface analysis defined specific residues whilst the RDC would define domain orientations, which is at the present time the most likely method to determine the full PDI domain configuration. This future work would prove an invaluable model which can be used to determine how the multi-domain organisation works in synergy during protein function.

REFERENCES

Allerhand, A., Doddrell, D., Glushko, V., Cochran, D. W., Wenkert, E., Lawson, P. J. and Gurd, F. R. (1971). Conformation and segmental motion of native and denatured ribonuclease A in solution. Application of natural-abundance carbon-13 partially relaxed Fourier transform nuclear magnetic resonance. *J Am Chem Soc*, **93** (2), 544-6.

Anelli, T., Alessio, M., Mezghrani, A., Simmen, T., Talamo, F., Bachi, A. and Sitia, R. (2002). ERp44, a novel endoplasmic reticulum folding assistant of the thioredoxin family. *Embo J*, **21** (4), 835-44.

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, **181** (96), 223-30.

Annunen, P., Helaakoski, T., Myllyharju, J., Veijola, J., Pihlajaniemi, T. and Kivirikko, K. I. (1997). Cloning of the human prolyl 4-hydroxylase alpha subunit isoform alpha(II) and characterization of the type II enzyme tetramer. The alpha(I) and alpha(II) subunits do not form a mixed alpha(I)alpha(II)beta2 tetramer. *J Biol Chem*, **272** (28), 17342-8.

Atzel, A. and Wetterau, J. R. (1993). Mechanism of microsomal triglyceride transfer protein catalyzed lipid transport. *Biochemistry*, **32** (39), 10444-50.

Atzel, A. and Wetterau, J. R. (1994). Identification of two classes of lipid molecule binding sites on the microsomal triglyceride transfer protein. *Biochemistry*, **33** (51), 15382-8.

Banhegyi, G., Lusini, L., Puskas, F., Rossi, R., Fulceri, R., Braun, L., Mile, V., di Simplicio, P., Mandl, J. and Benedetti, A. (1999). Preferential transport of glutathione versus glutathione disulfide in rat liver microsomal vesicles. *J Biol Chem*, **274** (18), 12213-6.

Bass, R., Ruddock, L. W., Klappa, P. and Freedman, R. B. (2004). A major fraction of endoplasmic reticulum-located glutathione is present as mixed disulfides with protein. *J Biol Chem*, **279** (7), 5257-62.

Bourdi, M., Demady, D., Martin, J. L., Jabbour, S. K., Martin, B. M., George, J. W. and Pohl, L. R. (1995). cDNA cloning and baculovirus expression of

the human liver endoplasmic reticulum P58: characterization as a protein disulfide isomerase isoform, but not as a protease or a carnitine acyltransferase. *Arch Biochem Biophys*, **323** (2), 397-403.

Brunger, A. T., Huber, R. and Karplus, M. (1987). Trypsinogen-trypsin transition: a molecular dynamics study of induced conformational change in the activation domain. *Biochemistry*, **26** (16), 5153-62.

Buchberger, A., Howard, M. J., Freund, S. M., Proctor, M., Butler, P. J., Fersht, A. R. and Bycroft, M. (2000). Biophysical characterization of elongin C from saccharomyces cerevisiae. *Biochemistry*, **39** (40), 12512.

Bukau, B. and Horwich, A. L. (1998). The Hsp70 and Hsp60 chaperone machines. *Cell*, **92** (3), 351-66.

Cabibbo, A., Pagani, M., Fabbri, M., Rocchi, M., Farmery, M. R., Bulleid, N. J. and Sitia, R. (2000). ERO1-L, a human protein that favors disulfide bond formation in the endoplasmic reticulum. *J Biol Chem*, **275** (7), 4827-33.

Chakravarthi, S., Jessop, C. E., Willer, M., Stirling, C. J. and Bulleid, N. J. (2007). Intracellular catalysis of disulfide bond formation by the human sulfhydryl oxidase, QSOX1. *Biochem J*, **404** (3), 403-11.

Chen, Y., Zhang, Y., Yin, Y., Gao, G., Li, S., Jiang, Y., Gu, X. and J., L. (2005). SPD--a web-based secreted protein database. *Nucleic acids research*, **33**, D169-D173.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T. J., Higgins, D. G. and Thompson, J. D. (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Res*, **31** (13), 3497-500.

Chivers, P. T. and Raines, R. T. (1997). General acid/base catalysis in the active site of Escherichia coli thioredoxin. *Biochemistry*, **36** (50), 15810-15816.

Clore, G. M., Driscoll, P. C., Wingfield, P. T. and Gronenborn, A. M. (1990). Analysis of the backbone dynamics of interleukin-1 beta using two-dimensional inverse detected heteronuclear 15N-1H NMR spectroscopy. *Biochemistry*, **29** (32), 7387-401.

Clore, G. M., Szabo, A., Bax, A., Kay, L. E., Driscoll, P. C. and Gronenborn, A. M. (1990). Deviations from the Simple 2-Parameter Model-Free Approach to the Interpretation of N-15 Nuclear Magnetic-Relaxation of Proteins. *Journal of the American Chemical Society*, **112** (12), 4989-4991.

Combet, C., Jambon, M., Deleage, G. and Geourjon, C. (2002). Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics*, **18** (1), 213-4.

Coppock, D. L., Kopman, C., Scandalis, S. and Gilleran, S. (1993). Preferential gene expression in quiescent human lung fibroblasts. *Cell Growth Differ*, **4** (6), 483-93.

Cornilescu, G., Delaglio, F. and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*, **13** (3), 289-302.

Creighton, T. E., Zapun, A. and Darby, N. J. (1995). Mechanisms and catalysts of disulfide bond formation in proteins. *Trends in biotechnology*, **13** (1), 18-23.

Cunnea, P. M., Miranda-Vizuete, A., Bertoli, G., Simmen, T., Damdimopoulos, A. E., Hermann, S., Leinonen, S., Huikko, M. P., Gustafsson, J. A., Sitia, R. and Spyrou, G. (2003). ERdj5, an endoplasmic reticulum (ER)-resident protein containing DnaJ and thioredoxin domains, is expressed in secretory cells or following ER stress. *J Biol Chem*, **278** (2), 1059-66.

Cuozzo, J. W. and Kaiser, C. A. (1999). Competition between glutathione and protein thiols for disulphide-bond formation. *Nat Cell Biol*, **1** (3), 130-5.

Darby, N. J. and Creighton, T. E. (1995). Functional properties of the individual thioredoxin-like domains of protein disulfide isomerase. *Biochemistry*, **34** (37), 11725-11735.

Darby, N. J., Freedman, R. B. and Creighton, T. E. (1994). Dissecting the mechanism of protein disulfide isomerase: catalysis of disulfide bond formation in a model peptide. *Biochemistry*, **33** (25), 7937-47.

Darby, N. J., Penka, E. and Vincentelli, R. (1998). The multi-domain structure of protein disulfide isomerase is essential for high catalytic efficiency. *Journal of Molecular biology*, **276** (1), 239-247.

Delorenzo, F., Goldberger, R. F., Steers, E. J., Givol, D. and Anfinsen, C. B. (1966). Purification and Properties of an Enzyme from Beef Liver Which Catalyzes Sulfhydryl-Disulfide Interchange in Proteins. *Journal of Biological Chemistry*, **241**, 1562-1567.

Denisov, A. Y., Kozlov, G., Gravel, M., Sprules, T., Braun, P. E. and Gehring, K. (2006). 1H, 13C and 15N resonance assignments of the catalytic domain of the goldfish RICH protein. *J Biomol NMR*, **36 Suppl 1**, 75.

Dyson, H. J., Jeng, M. F., Tennant, L. L., Slaby, I., Lindell, M., Cui, D. S., Kuprin, S. and A., H. (1997). Effects of buried charged groups on cysteine thiol ionization and reactivity in Escherichia coli thioredoxin: structural and functional characterization of mutants of Asp 26 and Lys 57. *Biochemistry*, **36** (9), 2622-2636.

Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A. and Rutter, W. J. (1985). Sequence of protein disulphide isomerase and implications of its relationship to thioredoxin. *Nature*, **317** (6034), 267-270.

Ellgaard, L. and Frickel, E. M. (2003a). Calnexin, calreticulin, and ERp57: teammates in glycoprotein folding. *Cell Biochem Biophys*, **39** (3), 223-47.

Ellgaard, L. and Helenius, A. (2001). ER quality control: towards an understanding at the molecular level. *Curr Opin Cell Biol*, **13** (4), 431-7.

Ellgaard, L. and Helenius, A. (2003b). Quality control in the endoplasmic reticulum. *Nat Rev Mol Cell Biol*, **4** (3), 181-91.

Englander, S. W. and Kallenbach, N. R. (1983). Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys*, **16** (4), 521-655.

Evans, J. N. S. (1995). Biomolecular NMR spectroscopy. (Oxford University Press).

Ferrari, D. M. and Soling, H. D. (1999). The protein disulphide-isomerase family: unravelling a string of folds. *Biochem J*, **339 (Pt 1)**, 1-10.

Fewell, S. W., Travers, K. J., Weissman, J. S. and Brodsky, J. L. (2001). The action of molecular chaperones in the early secretory pathway. *Annu Rev Genet*, **35**, 149-91.

Flynn, G. C., Pohl, J., Flocco, M. T. and Rothman, J. E. (1991). Peptide-binding specificity of the molecular chaperone BiP. *Nature*, **353** (6346), 726-30.

Forman-Kay, J. D., Clore, G. M., Wingfield, P. T. and Gronenborn, A. M. (1991). High-resolution three-dimensional structure of reduced recombinant human thioredoxin in solution. *Biochemistry*, **30** (10), 2685-2698.

Frand, A. R., Cuozzo, J. W. and Kaiser, C. A. (2000a). Pathways for protein disulphide bond formation. *Trends Cell Biol*, **10** (5), 203-10.

Frand, A. R. and Kaiser, C. A. (1998). The ERO1 gene of yeast is required for oxidation of protein dithiols in the endoplasmic reticulum. *Mol Cell*, **1** (2), 161-70.

Frand, A. R. and Kaiser, C. A. (1999). Ero1p oxidizes protein disulfide isomerase in a pathway for disulfide bond formation in the endoplasmic reticulum. *Mol Cell*, **4** (4), 469-77.

Frand, A. R. and Kaiser, C. A. (2000b). Two pairs of conserved cysteines are required for the oxidative activity of Ero1p in protein disulfide bond formation in the endoplasmic reticulum. *Mol Biol Cell*, **11** (9), 2833-43.

Freedman, R. B. (1984). Native disulphide bond formation in protein-biosynthesis - evidence for the role of protein disulphide isomerase. *Trends in Biochemical Science*, **9**, 438-441.

Freedman, R. B. (1995). The formation of protein disulphide bonds. *Current Opinion in Structural Biology* **5**(1), 85-91.

Freedman, R. B., Gane, P. J., Hawkins, H. C., Hlodan, R., McLaughlin, S. H. and Parry, J. W. (1998). Experimental and theoretical analyses of the domain architecture of mammalian protein disulphide-isomerase. *Biological Chemistry*, **379** (3), 321-328.

Freedman, R. B., Hirst, T. R. and Tuite, M. F. (1994). Protein disulphide isomerase: building bridges in protein folding. *Trends in biochemical sciences*, **19** (8), 331-336.

Freedman, R. B., Klappa, P. and Ruddock, L. W. (2002). Protein disulfide isomerases exploit synergy between catalytic and specific binding domains. *EMBO reports*, **3** (2), 136-140.

Frickel, E. M., Frei, P., Bouvier, M., Stafford, W. F., Helenius, A., Glockshuber, R. and Ellgaard, L. (2004). ERp57 is a multifunctional thiol-disulfide oxidoreductase. *J Biol Chem*, **279** (18), 18277-87.

Fuson, M. M. and Prestegard, J. H. (1983). Dynamics of an interfacial methylene in dimyristoylphosphatidylcholine vesicles using carbon-13 spin relaxation. *Biochemistry*, **22** (5), 1311-6.

Garbi, N., Tanaka, S., Momburg, F. and Hammerling, G. J. (2006). Impaired assembly of the major histocompatibility complex class I peptide-loading complex in mice deficient in the oxidoreductase ERp57. *Nat Immunol*, **7** (1), 93-102.

Geoghegan, K. F., Dixon, H. B. F., Rosner, P. J., Hoth, L. R., Lanzetti, A. J., Borzilleri, K. A., Marr, E. S., Pezullo, L. H., Martin, L. B., LeMotta, P. K., McColl, A. S., Kamth, A. V. and Stroh, J. G. (1999). Spontaneous -N-6-Phosphogluconoylation of a “His Tag” in Escherichia coli: The Cause of Extra Mass of 258 or 178 Da in Fusion Proteins. *Analytical Biochemistry*, **267**, 169-184.

Gilchrist, A., Au, C. E., Hiding, J., Bell, A. W., Fernandez-Rodriguez, J., Lesimple, S., Nagaya, H., Roy, L., Gosline, S. J., Hallett, M., Paiement, J., Kearney, R. E., Nilsson, T. and Bergeron, J. J. (2006). Quantitative proteomics analysis of the secretory pathway. *Cell*, **127** (6), 1265-81.

Gillece, P., Luz, J. M., Lennarz, W. J., de La Cruz, F. J. and Romisch, K. (1999). Export of a cysteine-free misfolded secretory protein from the endoplasmic reticulum for degradation requires interaction with protein disulfide isomerase. *J Cell Biol*, **147** (7), 1443-56.

Goldberger, R. F., Epstein, C. J. and Anfinsen, C. B. (1963). Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *J Biol Chem*, **238**, 628-35.

Grzesiek, S. and Bax, A. (1992a). Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.*, **114** (15), 6291-6293.

Grzesiek, S. and Bax, A. (1992b). An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins *J. Mag. Res.*, **99**, 201-207.

Hamman, B. D., Hendershot, L. M. and Johnson, A. E. (1998). BiP maintains the permeability barrier of the ER membrane by sealing the lumenal end of the translocon pore before and early in translocation. *Cell*, **92** (6), 747-58.

Hammes, G. G. (2005). Spectroscopy for the biological sciences. Wiley-InterScience, New York.

Hanahan, D. (1985). DNA cloning, a practical approach. IRL Press: Oxford.

Hatahet, F. and Ruddock, L. W. (2007). Substrate recognition by the protein disulfide isomerases. *FEBS journal*, **274** (20), 5223-52234.

Hebert, D. N., Foellmer, B. and Helenius, A. (1995). Glucose trimming and reglucosylation determine glycoprotein association with calnexin in the endoplasmic reticulum. *Cell*, **81** (3), 425-33.

Helenius, A. and Aebi, M. (2004). Roles of N-linked glycans in the endoplasmic reticulum. *Annu Rev Biochem*, **73**, 1019-1049.

Helenius, A., Trombetta, E. S., Hebert, D. N. and Simons, J. F. (1997). Calnexin, calreticulin and the folding of glycoproteins. *Trends in Cell Biology*, **7** (5), 193-200.

Hendershot, L., Wei, J., Gaut, J., Melnick, J., Aviel, S. and Argon, Y. (1996). Inhibition of immunoglobulin folding and secretion by dominant negative BiP ATPase mutants. *Proc Natl Acad Sci U S A*, **93** (11), 5269-74.

Hershko, A. and Ciechanover, A. (1998). The ubiquitin system. *Annu Rev Biochem*, **67**, 425-79.

Higgins, J. A. and Hutson, J. L. (1984). The roles of Golgi and endoplasmic reticulum in the synthesis and assembly of lipoprotein lipids in rat hepatocytes. *J Lipid Res*, **25** (12), 1295-305.

Hillson, D. A., Lambert, N. and Freedman, R. B. (1984). Formation and isomerization of disulfide bonds in proteins: protein disulfide-isomerase. *Methods Enzymology*, **107**, 281–294.

Hoober, K. L., Sheasley, S. L., Gilbert, H. F. and Thorpe, C. (1999). Sulfhydryl oxidase from egg white. A facile catalyst for disulfide bond formation in proteins and peptides. *J Biol Chem*, **274** (32), 22147-50.

Hosoda, A., Kimata, Y., Tsuru, A. and Kohno, K. (2003). JPDI, a novel endoplasmic reticulum-resident protein containing both a BiP-interacting J-domain and thioredoxin-like motifs. *J Biol Chem*, **278** (4), 2669-76.

Hwang, C., Sinskey, A. J. and Lodish, H. F. (1992). Oxidized redox state of glutathione in the endoplasmic reticulum. *Science*, **257** (5076), 1496-502.

John, D. C., Grant, M. E. and Bulleid, N. J. (1993). Cell-free synthesis and assembly of prolyl 4-hydroxylase: the role of the beta-subunit (PDI) in preventing misfolding and aggregation of the alpha-subunit. *Embo J*, **12** (4), 1587-95.

Kanai, S., Toh, H., Hayano, T. and M., K. (1998). Molecular evolution of the domain structures of protein disulfide isomerases. *Journal of molecular evolution*, **47** (2), 200-210.

Katti, S. K., LeMaster, D. M. and Eklund, H. (1990). Crystal structure of thioredoxin from Escherichia coli at 1.68 A resolution. *Journal of Molecular biology*, **212** (1), 167-184.

Kaufman, R. J. (2002). Orchestrating the unfolded protein response in health and disease. *J Clin Invest*, **110** (10), 1389-98.

Kawasaki, T. and Kasai, M. (1994). Regulation of calcium channel in sarcoplasmic reticulum by calsequestrin. *Biochem Biophys Res Commun*, **199** (3), 1120-7.

Kay, L. E., Torchia, D. A. and Bax, A. (1989). Backbone dynamics of proteins as studied by 15N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry*, **28** (23), 8972-9.

Kelly, S. M. and Price, N. C. (1997). The application of circular dichroism to studies of protein folding and unfolding. *Biochim Biophys Acta*, **1338** (2), 161-85.

Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. and Creighton, T. E. (1996). Structure determination of the N-terminal thioredoxin-like domain of protein disulfide isomerase using multidimensional heteronuclear 13C/15N NMR spectroscopy. *Biochemistry*, **35** (24), 7684-7691.

Kemmink, J., Darby, N. J., Dijkstra, K., Nilges, M. and Creighton, T. E. (1997). The folding catalyst protein disulfide isomerase is constructed of active and inactive thioredoxin modules. *Current biology*, **7** (4), 239-245.

Kemmink, J., Dijkstra, K., Mariani, M., Scheek, R. M., Penka, E., Nilges, M. and Darby, N. J. (1999). The structure in solution of the b domain of protein disulfide isomerase. *Journal of Biomolecular NMR*, **13** (4), 357-368.

Kemmink, J., Dijkstra, K., Mariani, M., Scheek, R. M., Penka, E., Nilges, M. and Darby, N. J. (1999). The structure in solution of the b domain of protein disulfide isomerase. *Journal of Biomolecular NMR*, **13** (4), 357-368.

Kivirikko, K. I. and Myllyharju, J. (1998). Prolyl 4-hydroxylases and their protein disulfide isomerase subunit. *Matrix Biol*, **16** (7), 357-68.

Kivirikko, K. I., Myllyla, R. and Pihlajaniemi, T. (1989). Protein hydroxylation: prolyl 4-hydroxylase, an enzyme with four cosubstrates and a multifunctional subunit. *Faseb J*, **3** (5), 1609-17.

Klappa, P., Ruddock, L. W., Darby, N. J. and Freedman, R. B. (1998). The b' domain provides the principal peptide-binding site of protein disulfide isomerase but all domains contribute to binding of misfolded proteins. *Embo J*, **17** (4), 927-35.

Koivu, J., Myllyla, R., Helaakoski, T., Pihlajaniemi, T., Tasanen, K. and Kivirikko, K. I. (1987). A single polypeptide acts both as the beta subunit of prolyl 4-hydroxylase and as a protein disulfide-isomerase. *J Biol Chem*, **262** (14), 6447-9.

Koivunen, P., Helaakoski, T., Annunen, P., Veijola, J., Raisanen, S., Pihlajaniemi, T. and Kivirikko, K. I. (1996). ERp60 does not substitute for

protein disulphide isomerase as the beta-subunit of prolyl 4-hydroxylase. *Biochem J*, **316 (Pt 2)**, 599-605.

Kornfeld, R. and Kornfeld, S. (1985). Assembly of asparagine-linked oligosaccharides. *Annu Rev Biochem*, **54**, 631-64.

Kozlov, G., Maattanen, P., Schrag, J. D., Pollock, S., Cygler, M., Nagar, B., Thomas, D. Y. and Gehring, K. (2006). Crystal Structure of the bb' Domains of the Protein Disulfide Isomerase ERp57. *Structure*, **14**, 1331–1339.

Laemmli, U. K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680-5.

Lamberg, A., Jauhiainen, M., Metso, J., Ehnholm, C., Shoulders, C., Scott, J., Pihlajaniemi, T. and Kivirikko, K. I. (1996). The role of protein disulphide isomerase in the microsomal triacylglycerol transfer protein does not reside in its isomerase activity. *Biochem J*, **315 (Pt 2)**, 533-6.

Lamberg, A., Pihlajaniemi, T. and Kivirikko, K. I. (1995). Site-directed mutagenesis of the alpha subunit of human prolyl 4-hydroxylase. Identification of three histidine residues critical for catalytic activity. *J Biol Chem*, **270** (17), 9926-31.

Land, A., Zonneveld, D. and Braakman, I. (2003). Folding of HIV-1 envelope glycoprotein involves extensive isomerization of disulfide bonds and conformation-dependent leader peptide cleavage. *Faseb J*, **17** (9), 1058-67.

Li, J., Zhang, S. and Wang, C. (2001). Effects of macromolecular crowding on the refolding of glucose- 6-phosphate dehydrogenase and protein disulfide isomerase. *J Biol Chem*, **276** (37), 34396-401.

Li, S. J., Hong, X. G., Shi, Y. Y., Li, H. and Wang, C. C. (2006). Annular arrangement and collaborative actions of four domains of protein-disulfide isomerase: a small angle X-ray scattering study in solution. *J Biol Chem*, **281** (10), 6581-8.

Liepinsh, E., Baryshev, M., Sharipo, A., Ingelman-Sundberg, M., Otting, G. and Mkrtchian, S. (2001). Thioredoxin fold as homodimerization module in the

putative chaperone ERp29: NMR structures of the domains and experimental model of the 51 kDa dimer. *Structure*, **9** (6), 457-71.

Linderstrom-Lang, K. (1955). Deuterium exchange between peptides and water. *Chem. Soc. (london) Spec. Publ.*, **2**, 1-20.

Lipari, G. and Szabo, A. (1981). Nuclear magnetic resonance relaxation in nucleic acid fragments: models for internal motion. *Biochemistry*, **20** (21), 6250-6.

Lipari, G. and Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity. *Journal of the American Chemical Society*, **104** (17), 4546-4559.

Lobley, A., Whitmore, L. and Wallace, B. A. (2002). DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics*, **18** (1), 211-2.

Lyles, M. M. and Gilbert, H. F. (1991). Catalysis of the oxidative folding of ribonuclease A by protein disulfide isomerase: pre-steady-state kinetics and the utilization of the oxidizing equivalents of the isomerase. *Biochemistry*, **30** (3), 619-25.

Lyles, M. M. and Gilbert, H. F. (1994). Mutations in the thioredoxin sites of protein disulfide isomerase reveal functional nonequivalence of the N- and C-terminal domains. *J Biol Chem*, **269** (49), 30946-52.

Macer, D. R. and Koch, G. L. (1988). Identification of a set of calcium-binding proteins in reticuloplasm, the luminal content of the endoplasmic reticulum. *Journal of cell science*, **91**, 61-70.

MacKay, J. P., Shaw, G. L. and King, G. F. (1996). Backbone dynamics of the c-Jun leucine zipper: 15N NMR relaxation studies. *Biochemistry*, **35** (15), 4867-77.

Martin, J. L. (1995). Thioredoxin-a fold for all reasons. *Structure*, **15** (3), 245-50.

Matthews, S. (2004). Perdeuteration/site-specific protonation approaches for high-molecular-weight proteins. *Methods Mol Biol*, **278**, 35-45.

Maudsley, A. A. and Ernst, R. R. (1977). Indirect detection of magnetic resonance by heteronuclear two dimensional spectroscopy. ***Chem. Phys. Lett.***, **50** (3), 368-372.

McArthur, A. G., Knodler, L. A., Silberman, J. D., Davids, B. J., Gillin, F. D. and Sogin, M. L. (2001). The evolutionary origins of eukaryotic protein disulfide isomerase domains: new evidence from the Amitochondriate protist Giardia lamblia. *Molecular biology and evolution*, **18** (8), 1455-63.

Miller, W. (2005). Minireview: regulation of steroidogenesis by electron transfer. *Endocrinology.*, **146** (6), 2544-2550.

Morimoto, R. I. (1998). Regulation of the heat shock transcriptional response: cross talk between a family of heat shock factors, molecular chaperones, and negative regulators. *Genes Dev*, **12** (24), 3788-96.

Morjana, N. A. and Gilbert, H. F. (1991). Effect of protein and peptide inhibitors on the activity of protein disulfide isomerase. *Biochemistry*, **30** (20), 4985-90.

Munro, S. and Pelham, H. R. (1987). A C-terminal signal prevents secretion of luminal ER proteins. *Cell*, **48** (5), 899-907.

Myllyharju, J. and Kivirikko, K. I. (1997). Characterization of the iron- and 2-oxoglutarate-binding sites of human prolyl 4-hydroxylase. *Embo J*, **16** (6), 1173-80.

Nakano, M., Maeno, A., Sasakawa, H., Yamaguchi, Y., Kikuchi, J., Asami, O., Kajino, T. and Kato, K. (to be published). Solution structure of the b' domain of thermophilic fungal protein disulfide isomerase .

Nakano, M., Murakami, C., Yamaguchi, Y., Sasakawa, H., Harada, T., Kurimoto, E., Asami, O., Kajino, T. and Kato, K. (2006). NMR assignments of the b' and a' domains of thermophilic fungal protein disulfide isomerase. *J Biomol NMR*, **36 Suppl 1**, 44.

Netzer, W. J. and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature*, **388** (6640), 343-9.

Nguyen, V. D., Wallis, K., Howard, M. J., Haapalainen, A. M., Salo, K. E. H., Saaranen, M. J., Sidhu, A., Wierenga, R. K., Freedman, R. B., Ruddock, L. W. and Williamson, R. A. ((Manuscript in preparation)). The b'x domain of human protein disufide isomerase can adopt alternative conformations: The structure of a mutant b'x identifies a novel interaction between the x region and the b'domain.

Oberdorf, J., Carlson, E. J. and Skach, W. R. (2001). Redundancy of mammalian proteasome beta subunit function during endoplasmic reticulum associated degradation. *Biochemistry*, **40** (44), 13397-405.

Otomo, T., Ito, N., Kyogoku, Y. and Yamazaki, T. (1999). NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. *Biochemistry*, **38** (49), 16040-4.

Pankalainen, M., Aro, H., Simons, K. and Kivirikko, K. I. (1970). Protocollagen proline hydroxylase: molecular weight, subunits and isoelectric point. *Biochim Biophys Acta*, **221** (3), 559-65.

Pariser, H. P., Rakeman, A. S. and Hausman, R. E. (1998). Thioreductase activity of retina cognin and its role in cell adhesion. *Brain research. Developmental brain research*, **11** (1), 1-9.

Parodi, A. J. (1991). Role of N-oligosaccharide endoplasmic reticulum processing reactions in glycoprotein folding and degradation. *The Biochemical journal*, **348** (1), 1-13.

Persson, S., Rosenquist, M., Knoblach, B., Khosravi-Far, R., Sommarin, M. and Michalak, M. (2005). Diversity of the protein disulfide isomerase family: identification of breast tumor induced Hag2 and Hag3 as novel members of the protein family. *Mol Phylogenet Evol*, **36** (3), 734-40.

Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997). Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci U S A*, **94** (23), 12366-71.

Phillips, J. L., Holdengreber, V., Ben-Shaul, Y., Zhang, J., Tolan, D. R. and Hausman, R. E. (1997). Developmental localization of retina cognin synthesis by in situ hybridization. *Brain research. Developmental brain research*, **104**, 143-152.

Pihlajaniemi, T., Helaakoski, T., Tasanen, K., Myllylä, R., Huhtala, M. L., Koivu, J. and Kivirikko, K. I. (1987). Molecular cloning of the beta-subunit of human prolyl 4-hydroxylase. This subunit and protein disulphide isomerase are products of the same gene. *EMBO journal*, **6** (3), 643-649.

Piotto, M., Saudek, V. and Sklenar, V. (1992). Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR*, **2** (6), 661-5.

Powis, G. and Montfort, W. R. (2001). Properties and biological activities of thioredoxins. *Annual review of biophysics and biomolecular structure* **30**, 421-55.

Raines, R. T. (1997). Nature's transitory covalent bond. *Nature structural biology*, **4** (6), 424-427.

Ramachandran, G. N., Ramakrishnan, C. and Sasisekharan, V. (1963). Stereochemistry of polypeptide chain configurations. *J Mol Biol*, **7**, 95-9.

Ricci, B., Sharp, D., O'Rourke, E., Kienzle, B., Blinderman, L., Gordon, D., Smith-Monroy, C., Robinson, G., Gregg, R. E., Rader, D. J. and et al. (1995). A 30-amino acid truncation of the microsomal triglyceride transfer protein large subunit disrupts its interaction with protein disulfide-isomerase and causes abetalipoproteinemia. *J Biol Chem*, **270** (24), 14281-5.

Ryabov, Y. and Fushman, D. (2006). Interdomain mobility in di-ubiquitin revealed by NMR. *Proteins*, **63** (4), 787-96.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. and Wuthrich, K. (1998). TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci U S A*, **95** (23), 13585-90.

Salzmann, M., Pervushin, K., Wider, G., Senn, H. and Wüthrich, K. (2000). NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. *J. Am. Chem. Soc.*, **122** (31), 7543-7548.

Schrag, J. D., Procopio, D. O., Cygler, M., Thomas, D. Y. and Bergeron, J. J. (2003). Lectin control of protein folding and sorting in the secretory pathway. *Trends Biochem Sci*, **28** (1), 49-57.

Schuck, P. (2000). Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J*, **78** (3), 1606-19.

Sevier, C. S. and Kaiser, C. A. (2002). Formation and transfer of disulphide bonds in living cells. *Nature reviews. Molecular cell biology*, **3** (11), 836-47.

Sevier, C. S. and Kaiser, C. A. (2008). Ero1 and redox homeostasis in the endoplasmic reticulum. *Biochimica et biophysica acta*, **(in press)**.

Sevier, C. S., Qu, H., Heldman, N., Gross, E., Fass, D. and Kaiser, C. A. (2007). Modulation of cellular disulfide-bond formation and the ER redox environment by feedback regulation of Ero1. *Cell*, **129** (2), 333-44.

Shoulders, C. C., Narcisi, T. M., Read, J., Chester, A., Brett, D. J., Scott, J., Anderson, T. A., Levitt, D. G. and Banaszak, L. J. (1994). The abetalipoproteinemia gene is a member of the vitellogenin family and encodes an alpha-helical domain. *Nat Struct Biol*, **1** (5), 285-6.

Silvennoinen, L., Karvonen, P., Koivunen, P., Myllyharju, J., Kivirikko, K. and Kilpelainen, I. (2001). Assignment of 1H, 13C and 15N resonances of the a' domain of ERp57. *J Biomol NMR*, **20** (4), 385-6.

Silvennoinen, L., Koivunen, P., Myllyharju, J., Kilpelainen, I. and Permi, P. (2005). NMR assignment of the N-terminal domain a of the glycoprotein chaperone ERp57. *J Biomol NMR*, **33** (2), 136.

Sitia, R. and Braakman, I. (2003). Quality control in the endoplasmic reticulum protein factory. *Nature*, **426** (6968), 891-4.

Smith, G. M., Yu, L. P. and Domingues, D. J. (1987). Directly observed 15N NMR spectra of uniformly enriched proteins. *Biochemistry*, **26** (8), 2202-7.

Solda, T., Garbi, N., Hammerling, G. J. and Molinari, M. (2006). Consequences of ERp57 deletion on oxidative folding of obligate and facultative clients of the calnexin cycle. *J Biol Chem*, **281** (10), 6219-26.

Song, J. L. and Wang, C. C. (1995). Chaperone-like activity of protein disulfide-isomerase in the refolding of rhodanese. *Eur J Biochem*, **231** (2), 312-6.

Spitzfaden, C., Weber, H. P., Braun, W., Kallen, J., Wider, G., Widmer, H., Walkinshaw, M. D. and Wuthrich, K. (1992). Cyclosporin A-cyclophilin complex formation. A model based on X-ray and NMR data. *FEBS Lett*, **300** (3), 291-300.

Sreerama, N., Venyaminov, S. Y. and Woody, R. W. (1999). Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci*, **8** (2), 370-80.

Stone, M. J., Fairbrother, W. J., Palmer, A. G., 3rd, Reizer, J., Saier, M. H., Jr. and Wright, P. E. (1992). Backbone dynamics of the Bacillus subtilis glucose permease IIA domain determined from 15N NMR relaxation measurements. *Biochemistry*, **31** (18), 4394-406.

Suh, J. K., Poulsen, L. L., Ziegler, D. M. and Robertus, J. D. (1999). Yeast flavin-containing monooxygenase generates oxidizing equivalents that control protein folding in the endoplasmic reticulum. *Proc Natl Acad Sci U S A*, **96** (6), 2687-91.

Sun, X. X., Dai, Y., Liu, H. P., Chen, S. M. and Wang, C. C. (2000). Contributions of protein disulfide isomerase domains to its chaperone activity. *Biochim Biophys Acta*, **1481** (1), 45-54.

Swift, L. L., Zhu, M. Y., Kakkad, B., Jovanovska, A., Neely, M. D., Valyi-Nagy, K., Roberts, R. L., Ong, D. E. and Jerome, W. G. (2003). Subcellular localization of microsomal triglyceride transfer protein. *J Lipid Res*, **44** (10), 1841-9.

Thorpe, C., Hoober, K. L., Raje, S., Glynn, N. M., Burnside, J., Turi, G. K. and Coppock, D. L. (2002). Sulfhydryl oxidases: emerging catalysts of protein

disulfide bond formation in eukaryotes. *Arch Biochem Biophys*, **405** (1), 1-12.

Tian, G., Xiang, S., Noiva, R., Lennarz, W. J. and H., S. (2006). The crystal structure of yeast protein disulfide isomerase suggests cooperativity between its active sites. *Cell*, **124** (1), 61-73.

Tochio, N., Koshiba, S., Inoue, M., Kigawa, T. and Yokoyama, S. (to be published). The solution structure of the second thioredoxin-like domain of human Protein disulfide-isomerase.

Tu, B. P., Ho-Schleyer, S. C., Travers, K. J. and Weissman, J. S. (2000). Biochemical basis of oxidative protein folding in the endoplasmic reticulum. *Science*, **290** (5496), 1571-4.

Tuderman, L., Kuutti, E. R. and Kivirikko, K. I. (1975). An affinity-column procedure using poly(L-proline) for the purification of prolyl hydroxylase. Purification of the enzyme from chick embryos. *Eur J Biochem*, **52** (1), 9-16.

Van Den Diepstraten, C., Papay, K., Bolender, Z., Brown, A. and Pickering, J. G. (2003). Cloning of a novel prolyl 4-hydroxylase subunit expressed in the fibrous cap of human atherosclerotic plaque. *Circulation*, **108** (5), 508-11.

van Lith, M., Karala, A. R., Bown, D., Gatehouse, J. A., Ruddock, L. W., Saunders, P. T. and Benham, A. M. (2007). A developmentally regulated chaperone complex for the endoplasmic reticulum of male haploid germ cells. *Mol Biol Cell*, **18** (8), 2795-804.

Veijola, J., Pihlajaniemi, T. and Kivirikko, K. I. (1996). Co-expression of the alpha subunit of human prolyl 4-hydroxylase with BiP polypeptide in insect cells leads to the formation of soluble and insoluble complexes. Soluble alpha-subunit-BiP complexes have no prolyl 4-hydroxylase activity. *Biochem J*, **315 (Pt 2)**, 613-8.

Venetianer, P. and Straub, F. B. (1963). Enzymic formation of the disulfide bridges of ribonuclease. *Acta physiologica Academiae Scientiarum Hungaricae*, **24**, 41-53.

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. and Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins*, **59** (4), 687-96.

Vuori, K., Pihlajaniemi, T., Marttila, M. and Kivirikko, K. I. (1992b). Characterization of the human prolyl 4-hydroxylase tetramer and its multifunctional protein disulfide-isomerase subunit synthesized in a baculovirus expression system. *Proc Natl Acad Sci U S A*, **89** (16), 7467-70.

Vuori, K., Pihlajaniemi, T., Myllyla, R. and Kivirikko, K. I. (1992a). Site-directed mutagenesis of human protein disulphide isomerase: effect on the assembly, activity and endoplasmic reticulum retention of human prolyl 4-hydroxylase in Spodoptera frugiperda insect cells. *Embo J*, **11** (11), 4213-7.

Wagner, G. (1993). Prospects for NMR of large proteins. *J Biomol NMR*, **3** (4), 375-85.

Walker, K. W., Lyles, M. M. and Gilbert, H. F. (1996). Catalysis of oxidative protein folding by mutants of protein disulfide isomerase with a single active-site cysteine. *Biochemistry*, **35** (6), 1972-80.

Wang, C. C. and Tsou, C. L. (1993). Protein disulfide isomerase is both an enzyme and a chaperone. *Faseb J*, **7** (15), 1515-7.

Wang, S., Trumble, W. R., Liao, H., Wesson, C. R., Dunker, A. K. and Kang, C. H. (1998). Crystal structure of calsequestrin from rabbit skeletal muscle sarcoplasmic reticulum. *Nature structural biology*, **5** (6), 476-483.

Weissman, J. S. and Kim, P. S. (1993). Efficient catalysis of disulphide bond rearrangements by protein disulphide isomerase. *Nature*, **365** (6442), 185-8.

Wetterau, J. R., Aggerbeck, L. P., Bouma, M. E., Eisenberg, C., Munck, A., Hermier, M., Schmitz, J., Gay, G., Rader, D. J. and Gregg, R. E. (1992). Absence of microsomal triglyceride transfer protein in individuals with abetalipoproteinemia. *Science*, **258** (5084), 999-1001.

Wetterau, J. R., Aggerbeck, L. P., Laplaud, P. M. and McLean, L. R. (1991). Structural properties of the microsomal triglyceride-transfer protein complex. *Biochemistry*, **30** (18), 4406-12.

Wetterau, J. R., Combs, K. A., McLean, L. R., Spinner, S. N. and Aggerbeck, L. P. (1991b). Protein disulfide isomerase appears necessary to maintain the catalytically active structure of the microsomal triglyceride transfer protein. *Biochemistry*, **30** (40), 9728-35.

Wetterau, J. R., Combs, K. A., Spinner, S. N. and Joiner, B. J. (1990). Protein disulfide isomerase is a component of the microsomal triglyceride transfer protein complex. *J Biol Chem*, **265** (17), 9800-7.

Wetterau, J. R., Lin, M. C. and Jamil, H. (1997). Microsomal triglyceride transfer protein. *Biochim Biophys Acta*, **1345** (2), 136-50.

Wetterau, J. R. and Zilversmit, D. B. (1984). A triglyceride and cholesteryl ester transfer protein associated with liver microsomes. *J Biol Chem*, **259** (17), 10863-6.

Wetterau, J. R. and Zilversmit, D. B. (1985). Purification and characterization of microsomal triglyceride and cholesteryl ester transfer protein from bovine liver microsomes. *Chem Phys Lipids*, **38** (1-2), 205-22.

Wetterau, J. R. and Zilversmit, D. B. (1986). Localization of intracellular triacylglycerol and cholesteryl ester transfer activity in rat tissues. *Biochim Biophys Acta*, **875** (3), 610-7.

Wiertz, E. J., Tortorella, D., Bogyo, M., Yu, J., Mothes, W., Jones, T. R., Rapoport, T. A. and Ploegh, H. L. (1996). Sec61-mediated transfer of a membrane protein from the endoplasmic reticulum to the proteasome for destruction. *Nature*, **384** (6608), 432-8.

Winter, J., Klappa, P., Freedman, R. B., Lilie, H. and Rudolph, R. (2002). Catalytic activity and chaperone function of human protein-disulfide isomerase are required for the efficient refolding of proinsulin. *J Biol Chem*, **277** (1), 310-7.

Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S. and Sykes, B. D. (1995a). 1H, 13C and 15N random coil NMR chemical shifts of the common amino

acids. I. Investigations of nearest-neighbor effects. *J Biomol NMR*, **5** (1), 67-81.

Wishart, D. S. and Sykes, B. D. (1994). The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. *J Biomol NMR*, **4** (2), 171-80.

Woodward, C., Simon, I. and Tuchsen, E. (1982). Hydrogen exchange and the dynamic structure of proteins. *Mol Cell Biochem*, **48** (3), 135-60.

Wuthrich, K., Wagner, G., Richarz, R. and Braun, W. (1980). Correlations between internal mobility and stability of globular proteins. *Biophys J*, **32** (1), 549-60.

Zapun, A., Darby, N. J., Tessier, D. C., Michalak, M., Bergeron, J. J. and Thomas, D. Y. (1998). Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. *J Biol Chem*, **273** (11), 6009-12.

Zapun, A., Petrescu, S. M., Rudd, P. M., Dwek, R. A., Thomas, D. Y. and Bergeron, J. J. (1997). Conformation-independent binding of monoglucosylated ribonuclease B to calnexin. *Cell*, **88** (1), 29-38.

Ziegler, D. M. and Poulsen, L. L. (1977). Protein disulfide bond synthesis: a possible intracellular mechanism. *Trends in Biochemical Sciences*, **2** (4), 79-81.

APPENDIX 1 – MANUSCRIPT IN PREPARATION

THE B'X DOMAIN OF HUMAN PROTEIN DISULPHIDE ISOMERASE CAN ADOPT ALTERNATIVE CONFORMATIONS: THE STRUCTURE OF A MUTANT B'X IDENTIFIES A NOVEL INTERACTION BETWEEN THE X REGION AND THE B' DOMAIN.

Nguyen Van Dat^§, Katrine Wallis*§, Mark J. Howard^^, Antti M. Haapalainen^, Kirsi E.H. Salo^, Mirva J. Saaranen^, Ateesh Sidhu*, Rik K. Wierenga^, Robert B Freedman*, Lloyd W. Ruddock^ and Richard A. Williamson^^

^ Department of Biochemistry, University of Oulu, FIN-90014 Oulu, FINLAND

* Department of Biological Sciences, Warwick University, Coventry CV4 7AL, UK

^^ Department of Biosciences, University of Kent, C

anterbury CT2 7NJ, UK

§ These authors contributed equally to this work

Running title: Interaction of the **x** linker with the **b'** domain of human PDI

Address correspondence to: Robert B. Freedman, Department of Biological Sciences, Warwick University, Gibbet Hill Road, Coventry CV4 7AL, UK. Tel.: +44-24-76523516; Fax: +44-24-76523568; E-mail: R.B.Freedman@warwick.ac.uk

Protein disulphide-isomerase (PDI) is a key protein folding catalyst in the endoplasmic reticulum and the b' domain of human PDI is essential for non-covalent binding of incompletely folded substrates. No high resolution crystal or NMR structure of the b' domain of human PDI has been determined. We have now shown by fluorescence and NMR that recombinant human PDI b'x (comprising the b' domain and the subsequent x linker region) can assume at least two different conformations in solution. We have resolved two conformers and shown that one is a monomer in which the unique Trp in the x region is immobilized in a hydrophobic environment, whereas the other is mainly a dimer in which the Trp is dynamic and exposed to the aqueous environment. We have screened mutants in the b'x region to identify mutations which favour one of these conformers and have isolated and characterised examples of both types. We have crystallised one variant (I272A mutation), identified as stabilizing the monomeric species, and determined its crystal structure to a resolution of 2.2 Å. This confirms that the

b' domain has the typical thioredoxin fold and that the x region can interact with the b' domain to bury the unique Trp by 'capping' a hydrophobic site on the b' domain. The position and orientation of the x linker sequence in this structure is different from that in yeast Pdi1p and we infer that the x region of PDI can adopt alternative conformations during the functional cycle of PDI action.

The protein disulphide-isomerases (PDIs [1]) are a family of proteins initially defined by their ability to assist the formation of native disulphide bonds in secretory and cell-surface proteins during their biosynthesis, but now characterised structurally by the presence of one or more thioredoxin-fold (trx) domains, and by their location within the lumen of the endoplasmic reticulum. The human PDI family is large (> 17 members) and the structures and specific roles of the multiple family members have not yet been defined (1); the best-characterised is PDI itself, the archetype protein folding catalyst. PDI from a number of mammalian species has been studied for many years (2) and its properties as a catalyst of disulphide-interchange and associated protein folding are well-understood (3-7).

When mammalian PDI was first cloned and sequenced, it was immediately recognised that it comprised multiple homologous domains (8) and subsequently the overall domain organization was defined as **a-b-b'-x-a'-c**, where **a** and **a'** are trx domains containing the active site dithiol sequence –WCGHC-, **b** and **b'** have the trx conformation but lack the active-site motif, **x** is a linker region and **c** is a C-terminal acidic tail (9-12). This domain organization was finally confirmed by the x-ray structural analysis of full-length yeast Pdi1p (13-14). While the **a** and **a'** domains of human PDI can catalyse simple thiol:disulphide interchange reactions, the presence of the **b'** domain is essential for catalysis of protein disulphide isomerizations and for non-covalent ligand binding (15-16); the domains operate synergistically in linking catalytic and chaperone activities in protein folding (10).

The **b'** domains of human PDI family members show greater sequence diversity than do the **b**, **a** and **a'** domains, and studies with hybrid and chimaeric species indicate that the specificities of PDIs for ligands and partner proteins are primarily defined by their **b'** domains; this has been confirmed in studies on the pancreas-specific isoform PDIp (17), on the PDI homologue ERp57, which interacts specifically with the lectins calnexin and calreticulin (18-21), and on the newly-described species ERp27 (22) which comprises only **b** and **b'**-like domains. These results make it of particular interest to study -- in molecular detail -- the **b'** domains of PDIs and their interactions with ligands. Frustratingly, the **b'** domain of human PDI is the only domain for which the structure has

not been solved and for several members of the family (e.g. PDIp, ERp27), the **b'** domains do not give good yields of well-folded soluble protein when expressed as recombinant constructs in *E.coli.*

We previously reported the expression and characterization of the **b'** and **b'x** domain fragments of human PDI (12). In this paper we report further characterization of the recombinant **b'x** fragment and show that it exists in two distinct conformations which differ in tertiary structure -- specifically in the environment of the unique Trp residue in the **x** region -- and in tendency to dimerize. In addition, we describe mutants of **b'x** which exclusively adopt one or other of these alternative conformations and we report the crystal structure of one of these mutants. The properties of the two conformers, and of the interconversion between them, may throw light on the dynamic interactions between the **b'** and **x** regions in full-length PDI. Furthermore, they may explain some of the difficulties in characterizing the molecular properties of PDI and its sub-fragments and in understanding the ability of PDI to interact with folding substrates.

EXPERIMENTAL PROCEDURES

Mutagenesis, expression and purification. The plasmid containing the gene encoding the human PDI **b'x** domain fragment is a derivative of pET23b (Novagen). The resulting protein is expressed with an N-terminal His tag MHHHHHHM and encodes residues K213 to P351 of mature PDI. Mutagenesis, transformation, expression and initial purification were done as described previously (12).

To subfractionate the **b'x** conformers, the eluate from initial purification on chelating Sepharose was dialysed overnight against 20 mM phosphate buffer, pH 7.3 and applied to a 10 ml Source30Q ion exchange column. The protein was eluted from the ion exchange column with a linear gradient from 0-250 mM NaCl over 100 ml. Fractions (2 ml) were pooled based on reducing native-PAGE and SDS-PAGE gels. Pools were concentrated using Vivaspin 20 columns with a molecular weight cut-off of 5000 Da. To obtain preparations enriched in monomeric and dimeric species, pools were applied to a Superdex 75 column equilibrated with 20 mM sodium phosphate, 150 mM NaCl pH 7.1.

Purification on a micro-scale from 5-10ml of bacterial culture for fluorescence screening of mutants was done using Ni-NTA spin columns (Qiagen) according to the protocol recommended by the manufacturer.

Spectroscopy. Fluorescence spectra were measured on a Photon Technology International fluorometer or on a Perkin/Elmer LS50 fluorometer. CD spectra were measured on a Jasco J-815 or J-715 CD spectrometer.

Analytical gel-filtration. Analytical gel-filtration was carried out on a Superdex 75 10/300 column (GE Healthcare) in 20 mM phosphate, 150 mM NaCl, pH 7.1. The markers used were horse heart cytochrome c (1.64 nm), bovine carbonic anhydrase (2.39 nm), bovine serum albumin (3.52 nm), yeast alcohol dehydrogenase (3.72 nm) and sweet potato α-amylase (4.15 nm); the values in brackets are the Stokes radii. $[-\log(K_{av})]^{\frac{1}{2}}$ was plotted against the Stokes radii of the marker proteins, where K_{av} is the partition coefficient determined as $(V_e - V_0)/(V_t - V_0)$, where V_e is the elution volume, V_t the total column volume and V_0 the void volume. This graph was then used to determine the Stokes radii.

Analytical ultracentrifugation. Analytical ultracentrifugation experiments and data analyses were as described in Girija et al (23). The protein samples were dialysed overnight into 20 mM phosphate, 150 mM NaCl, pH 7.1. The partial specific volume of 0.746 cm^3/g (based on the amino acid composition of the recombinant **b'x** fragment) was used in the calculation of molecular weight from sedimentation velocity data.

NMR spectroscopy. Cells were grown in minimal medium to allow isotopic labeling using 1g/l ^{15}N ammonium sulphate as sole nitrogen source and labelled protein was expressed and purified as above. Samples for NMR analysis were concentrated to between 0.36mM and 1mM in 20mM phosphate buffer (pH 6.5) containing 150mM NaCl by centrifugation on Vivaspin columns with a 5kD cut-off and then D_2O added to a final concentration of 10% (v/v). $^{15}N/^{1}H$ HSQC spectra were collected as described in Alanen et al (22).

Crystallization. To prepare selenium-labelled protein, a fresh colony of *E.*coli B834(DE3) carrying plasmid pAKL73 encoding for human PDI **b'x** I272A was inoculated into 200 ml of sterile M9 minimal media with an additional 0.5 mM L-methionine at 37°C, 200 rpm for 20 hours. The bacteria were collected by centifugation and washed twice with 200 ml of sterile water. The bacteria were then resuspended in M9 media with an additional 0.5 mM of selenonium-methionine to an OD_{600} of 0.3 and grown at 37°C, 200 rpm to an OD_{600} of 0.7 and then induced with 1 mM of IPTG at 25°C, 200rpm, for 12 hours. The bacteria were then harvested by centifugation, resuspended in 1/10 volume 20 mM phosphate pH 7.3, DNAase added to final concentration of 1 μg/l and lysozyme to final concentration of 1 mg/ml and then frozen. The protein was purified as for non-labelled protein. 10 mg of purified Se-labeled protein was obtained from 1 litre of culture.

Crystals of the SeMet derivative were obtained after one week by the hanging-drop, vapour diffusion equilibration method (24)

using 2μl of protein solution (10mg/ml) and 2μl well solution (0.2 M NaCl, 2.95 M ammonium sulfate, 0.1 M Tris-HCl buffer pH 9.5) at 22°C. Crystals of unlabelled protein were obtained by the same method at pH 7.5, 8.5 and 9.5. Crystals were harvested from drops, put into paraffin oil for a few seconds and then frozen in liquid nitrogen. X-ray data were collected at cryogenic temperature (100 K).

Data collection, structure determination and refinement. Multiwavelength anomalous dispersion (MAD) data were collected for the Se-Met crystal on CCD detector at beamline BW7A (DESY, Hamburg, Germany). The X-ray data from native crystals at pH7.5, 8.5 and 9.5 were collected at X12 (DESY, Hamburg, Germany). Images were processed using the XDS program package (25-26). Using the Auto-RickShaw protocol (27), the MAD-dataset resulted in the initial model that was completed with COOT (28). Subsequently the structure of the native datasets were refined using REFMAC5 (29) with the translation libration screw (TLS) description of the anisotropic rigid body motion (30). The data collection statistics and refinement statistics of the pH 8.5 structure are given in table 1. The model includes 49 waters and one sulphate ion.

RESULTS

*Preparations of recombinant human PDI **b'x** contain at least two conformers.* The domain boundaries of human PDI were defined in previous work (11-12); the constructs studied here correspond to the complete **b'** domain and subsequent **x** region with a short N-terminal tag. The **b'** domain of human PDI contains no Trp residues, but there is a single Trp residue located in the **x** region (W347). Despite only having a single tryptophan, preparations of unfractionated recombinant **b'x** gave a fluorescence emission spectrum with two maxima at c. 334 nm and 355 nm (Fig 1a). Furthermore, in HSQC spectra of ^{15}N-labelled human PDI **b'x** (Fig 1b) there were two peaks for the characteristic Trp indole sidechain resonance (shown boxed). There was, however, no indication of heterogeneity by mass spectrometry (data not shown). We have previously shown that **b'x** preparations give a CD spectrum consistent with substantial α-helix and β-sheet content and a single denaturation transition curve with a midpoint of 2.32M guanidine hydrochloride (12), both suggesting that the preparations do not contain any significant proportion of denatured protein. These results therefore indicated the possibility of conformational heterogeneity in the **b'x** preparations. Analysis by fluoresence spectroscopy and CD indicated that the ratio between these conformers could be altered by the applied salt concentration and that in dilute conditions

at room temperature this salt-induced change in structure was reversible (data not shown).

To better study the alternative conformers we attempted to resolve them by ion-exchange chromatography (Fig 2a). Fractions corresponding to distinct elution peaks were pooled and analysed. Electrospray MS of the pools showed that they each comprised a single protein species which had the identical mass of 17282 Da (expected 17283 Da); SDS-PAGE also indicated a single homogeneous species of c.17 kDa (see also Fig 4a, below). The pools were stored under various conditions in 150 mM NaCl, 20 mM phosphate buffer pH 7.3 and then re-analysed by ion-exchange chromatography. As shown in Figure 2b, the pool of early-eluting material was converted to a mixture of 'early' and 'late' eluting material after being stored frozen and then thawed, whereas similar material stored at 4°C showed much less conversion; the pool of 'late' eluting material was not affected significantly by storage (Fig 2c).

Spectroscopic analysis of the 'early' and 'late' pools confirmed that they differed in conformation. Both forms gave CD spectra characteristic of globular proteins with defined secondary structure, although the spectra were clearly different in detail (Fig 3a). The forms also differed markedly in fluorescence emission spectra (not shown); the 'early' pool showed a single broad emission peak with a maximum at 332 nm, characteristic of a Trp residue in a hydrophobic environment, while the 'late' pool showed a very broad maximum over the range 340-350nm, the latter being characteristic of a Trp residue in an exposed aqueous environment, plus a distinct peak with a maximum at 304nm, indicating inefficient energy transfer from Tyr residues in the protein to the single Trp. To confirm that neither pool represented a denatured form of the protein, pools of 'early' and 'late' fractions were titrated with guanidine hydrochloride to assess their respective stabilities to denaturation. Both species showed some changes in fluorescence at low concentrations of denaturant (< 0.5M) followed by a plateau in the concentration range 0.5-1.5 M and then a pronounced denaturation transition over the concentration range 1.5-2.5M with the emission maximum of the denatured protein being 350nm for both pools; the mid-points of these transitions were indistinguishable (Fig 3b). HSQC spectra of the two pools showed marked differences in peak resolution and dispersion; the 'early' pool (Fig 3c) gave a well resolved spectrum consistent with a folded protein of the expected molecular size, whereas the spectrum for the 'late' pool (Fig 3d) was much broader with fewer peaks discernible. This type of spectrum suggests a protein in conformation exchange with the possibility of some oligomer formation. The spectrum of the 'late' pool contains many peaks of the 'early' form of the protein showing that the sample is not pure. The

peaks from the 'late' form do not appear so obviously in the 'early' pool sample as they are much broader and therefore more difficult to see. The indole Trp crosspeak was predominantly in the downfield form in the 'early' pool sample (suggesting that this peak is from the indole sidechain experiencing a hydrophobic environment) and in the upfield form in the 'late' pool (suggesting this peak is from the indole in a more aqueous environment). It is interesting that the 'late' pool peak is sharper than that in the 'early' sample suggesting that the Trp residue in this material was in a region of polypeptide that was tumbling faster, and that its chemical shift matches that expected for a random coil (31).

Based on these results, we infer that the **b'x** fragment can exist in at least two alternative structured conformations. In one conformation (which we term the 'capped' form), the Trp residue in the **x** linker is buried in a hydrophobic environment sufficiently close to the Tyr residues of the **b'** domain for efficient Forster resonance energy transfer, whereas in the other conformation (which we term the 'uncapped' form) the Trp residue is fully exposed to solvent, remote from the Tyr residues of the **b'** domain, and showing dynamic behaviour indicative of being in a flexible region of polypeptide. Our results show that the 'capped' form predominates in early fractions from the ion-exchange column while the 'uncapped' form predominates in 'late' fractions.

*The 'capped' form of PDI **b'x** is a monomer in solution but the 'uncapped' form has a strong tendency to form dimers.* The marked differences in properties of the 'early' and 'late' fractions of wild-type PDI **b'x**, as seen by NMR (Fig 3c,d), led us to examine the oligomer status of the two pools. Analysis of ion-exchange fractions by SDS-PAGE showed no differences across the ion-exchange peaks but native polyacrylamide gel electrophoresis showed that later elution fractions showed both monomer and dimer bands, whereas the earlier fractions showed only a monomer band (Fig 4a). Analytical gel-filtration was carried out on 'early' and 'late' pools (Fig 4b) and showed clearly that both pools contained a mixture of two species differing in size, but that material of lower Stokes radius predominated in the 'early' pool and material of higher Stokes radius predominated in the 'late' pool. Based on markers, the Stokes radii for the two species were calculated as 2.12 nm and 2.68 nm respectively. The early pool sample was also studied by sedimentation velocity ultracentrifugation (not shown) and the data analysed using the experimental Stokes radii from the gel-filtration experiments. This confirmed the presence of two distinct species; a major species of mass 16.3 kDa, (corresponding to a monomer) and a minor contaminating species of mass 34.7 kDa (corresponding to a dimer).

Screening mutations that trap a single conformation of PDI ***b'x***. Since neither the 'early' nor 'late' pools of wild-type **b'x** represent homogenous preparations of 'capped' or 'uncapped' conformer, we decided to screen mutants in the **b'x** construct which might alter the conformer distribution and perhaps allow a single species to be characterised in detail. Since the fluorescence spectrum gave an easily-diagnosable read-out of the proportions of the 'capped' and 'uncapped' forms, this was chosen for the screen. In total 75 mutants of **b'x** were screened, a significant proportion of which were generated previously (12). These mutants were purified on a micro-scale and their fluorescence spectra recorded in the elution buffer immediately after purification. Many of the mutants -- like the wild type protein -- showed a mixed fluorescence spectrum with peaks c. 334nm and 355nm. However, some mutants showed a significant shift in fluorescence spectrum towards a single 355nm peak while others showed a significant shift towards a single 334nm peak (Fig 5a,b). Since the concentrations of the proteins obtained from the micro-scale purification varied by about three-fold, the parameter chosen to cross-compare spectra was the ratio of the average fluorescence at 331-337nm to the average fluorescence at 352-358nm. This ratio in elution buffer was 1.03 for the wild type protein and varied from 0.76 to 1.29 for the mutant proteins (Fig 5c). Many of the mutations which resulted in a significant shift in the fluorescence ratio were either located around the substrate binding site as defined previously (12) or were located in **x**.

Two red-shifted and two blue-shifted mutants were chosen for purification to homogeneity on the milligram scale. Fluorescence analysis of these mutant proteins confirmed the results obtained from the micro-scale purification (data not shown). The CD spectra were characteristic of globular proteins with defined secondary structure, although the spectra were clearly different in detail from each other and from the unfractionated wild-type protein (data not shown). The mutant proteins were generated in ^{15}N-labelled form and HSQC spectra of the blue-shifted I272A (Fig 5d) and D346A/D348A mutants (not shown) were very similar in terms of line widths and peak numbers to that of the 'capped' wild-type species (Fig 3c). In contrast, the HSQC spectra of the red-shifted L343A (Fig 5e) and K349A mutants (not shown) revealed that both purified proteins showed the poorly defined broad peaks characteristic of the 'uncapped' conformer of the wild-type protein. The proportion of 'uncapped' and 'capped' forms, as judged by the two crosspeaks from the Trp indole HN, showed that the two red-shifted mutants were entirely in the 'uncapped' form, whereas all blue-shifted mutants had substantially less of

the 'uncapped' form than the unfractionated wild-type sample (Fig 1b); the amount of 'uncapped' form in the blue-shifted mutants decreased in the order D346A/D348A > I272A > the triple mutant I272A/D346A/D348A, with the triple mutant containing essentially none of this conformer (not shown). These results suggest that the blue-shift mutations act to stabilise the 'capped' form relative to the 'uncapped' form and that the effect of the blue-shift mutations was additive.

*Crystallography of mutant PDI **b'x***. Despite numerous attempts, we have been unable to obtain crystals from wild-type PDI **b'** or **b'x**, a situation we tentatively assigned to the mixed conformers present. To test this hypothesis, we tried to crystallize the I272A and D346A/D348A mutants of human PDI **b'x** which had appeared by NMR to be predominantly in a single conformation. Both mutants gave crystals, but despite extensive optimization we were unable to improve the crystallization of the D346A/D348A mutant to obtain crystals which diffracted well. In contrast, we obtained crystals of the I272A mutant which diffracted well using 3 different conditions. Excluding the high B-factor loop which contains E304, E305 and E306 the crystals obtained under the three conditions gave structures whose backbone traces were effectively super-imposable. We chose to refine the structure of the pH 8.5 crystal, since the diffraction pattern was slightly better than that obtained at pH 7.5 and since it is closer to physiological pH than the crystal obtained at pH 9.5. This structure has been refined at 2.2Å resolution (table 1). There is one monomer per asymmetric unit. The final model contains 2 *cis*-peptide bonds (C295-P296 and E304-E305). Altogether seventeen residues were not observed in the electron density map, these being the three C-terminal residues and fourteen N-terminal residues, including the eight residues of the His-tag. Therefore the ordered part of the **b'x** domain extends from L219 to D348.

The crystal structure of the **b'** domain of the I272A **b'x** construct (Fig 6a,b) unsurprisingly exhibits a PDI-type trx fold (ßaßaßaßßa) with part of **x** (residues P336-L338) forming an additional β-strand to the mixed β-sheet core of the trx fold and the rest of **x** looping round to remain juxtaposed to **b'** with residues E345-D348 forming a short C-terminal helix.

The **b'** domain of PDI binds peptides and non-native proteins via predominantly hydrophobic interactions (16,17) and hence a hydrophobic surface or solvent exposed hydrophobic pocket might be expected on **b'**. Surface charge density maps (Fig 6c, d) of **b'** reveal a hydrophobic pocket on one face of the domain which is capped by the **x** region. Specifically, the side-chain of M339 (the first residue after the β-strand in the **x** region) binds in a hydrophobic pocket comprising the sidechains of F287, F288, I301, M307 and the

alkyl chain of K309. The side-chains of L343 and W347 (further along the **x** region) also make hydrophobic interactions with the **b'** domain and interact with the side-chains of F223, A228, P229, F232, I284, F287, F288, L303, M307 in a broad site immediately adjacent to the binding pocket of M339. The residues F223, I284, F287 and F288 interact most closely with the side-chain of W347, and the site of mutation (A272) is located behind F223 which forms the bottom of this interaction pocket. The indole HN of W347 makes polar interactions with the side-chain carboxyl of E225 which also forms a salt bridge with the guanidino group of R283 on the top rim of the site (Fig 6d). It is noteworthy that the residues M339, L343 and W347 in the **x** region, and those forming the hydrophobic patches on the **b'** domain with which they interact, are well conserved in PDI sequences of metazoan animal species.

This hydrophobic location of the side-chain of W347 is consistent with the single peaked blue-shifted fluorescence spectrum of this mutant. Structure analysis by ICM (Molsoft L.L.C., La Jolla, CA) shows a vacancy in the structure with a size of 50 $Å^3$ adjacent to the A272 side chain. This is consistent with the difference in volume between the side chain of the isoleucine found in the wild-type protein at this position and the alanine found in the mutant, suggesting that the structure obtained for the mutant is valid for that of the wild-type protein, or at least for the 'capped' conformer of it. This conclusion is strengthened by comparison of the HSQC data for the I272A mutant (Fig 5e) and for the wt 'capped' conformer (Fig 3c) which revealed that c. 80% of backbone amide cross-peaks were unperturbed by the residue substitution.

DISCUSSION

*A high resolution structure of the **b'x** region of human PDI.* The biological significance of mammalian PDI in the folding of disulphide-bonded secretory proteins has been clear for many years (2) and throughout that time pure protein has been abundantly available, either from mammalian tissues or recombinant sources (32-33). No high-resolution structure of a full-length PDI from a multicellular organism has been achieved to date, despite the fact that many groups have attempted to determine the structure of PDIs by x-ray crystallography. This may reflect inter-domain flexibility or the existence of regions within the full-length molecule that can adopt alternative conformations. The structures of individual **a**, **b** and **a'** domains of human PDI have been determined by NMR, but it is interesting that no structure has been achieved by this approach for the **b'** domain or the combination of **b'** domain and adjacent **x** linker region. The difficulty in previous studies of the **b'** domain or **b'x** combination could have arisen from poor definition of the domain

boundary, but it is possible that this domain does show some intrinsic flexibility or conformational heterogeneity.

We have now conducted NMR studies on recombinant preparations corresponding to the **b'**, **b'x**, **bb'** and **bb'x** domain combinations and have completed backbone assignments of both **b'x** and **bb'x** (manuscript in preparation). In the current paper we have analysed recombinant human **b'x** in detail -- using a wide range of techniques for structural study in solution -- and identified the existence of at least two conformational states which clearly differ in the structural relationship between the **x** region (specifically W347) and the **b'** domain. Wild-type **b'x** preparations can be sub-fractionated to generate samples which comprise mainly a monomeric species in which the environment of W347 can be characterised by fluorescence and NMR as being hydrophobic and immobilized; other sub-fractions comprise mainly a dimeric species in which the W347 residue is mobile and experiences an aqueous environment. Samples containing mainly these two alternative conformations are stable for periods of weeks in solution. We have not fully investigated the conditions under which they form or interconvert, but they are clearly both present at an early stage in our preparations.

In order to overcome the complexities introduced by the existence of these states, we sought to generate mutants which might preferentially stabilize one state and eliminate the conformational heterogeneity. Using fluorescence screening we were able to identify mutants which preferentially adopted one or other conformation and we have crystallized and determined the structure of the I272A mutant of **b'x**, which we inferred to be entirely in the 'capped' conformation. In accordance with our inference from solution studies on the wild-type **b'x**, the **x** region is intimately associated with the **b'** domain in the crystal structure of **b'x** I272A, and the indole side-chain of W347 is buried in a hydrophobic pocket. Most of the strong 'red-shift' mutants involve mutations in residues directly involved in the contact between **x** and **b'**. While many of the 'blue-shift' mutants are also in residues involved in this contact, others – including I272A and all of the mutations previously identified to inhibit substrate binding (12) – are not contact residues and we infer that these act indirectly by stabilizing the conformer in which the hydrophobic pocket is 'capped' by **x** relative to the 'uncapped' conformer.

*Comparison with **b'** and **x** conformations in other PDI family structures.* In Fig 7a we show a structure-based sequence alignment of the human PDI **b'x** region with the corresponding region of yeast Pdi1p, based on the structure of I272A **b'x** solved here and on the full-length structure of Pdi1p (13). This shows that, overall, the two **b'** domains adopt very similar

conformations with the exception of the region of Pdi1p (residues 330-344 in mature Pdi1p numbering) which forms a loop and a short additional helix. In human PDI **b'x**, this region comprises simply a short loop linking β5 directly to α4. A superposition of the backbone atoms (N, Cα, C) for the core β-strands (42 residues) of yeast and human **b'x** using the alignment in Fig 7a gave an RMSD of 0.6Å, this value increased to 1.5Å when both β-strands and helices were included (102 residues). Comparison using DALI (34) of human **b'x** to all structures in the protein databank other than yeast Pdi1p found closest similarity to the trx domains in (i) the **b'** domain of PDI from the thermophilic fungus *Humicola insolens* (2djk), (ii) *Drosophila* Wind chaperone (1onn and 2c0e), (iii) thioredoxin-2 from *Anabaena* (1thx), (iv) the C73S mutant of human thioredoxin (1erv) (v) domain 2 of rabbit skeletal calsequestrin (1a8y), (vi) the **b'** domain of human **bb'** fragment from Erp57 (2h8l) and (vii) the **b** domain of human PDI (1bjx).

However there is a major difference between the human and yeast PDI structures in the orientation of the **x** region, which in the Pdi1p structure does not cap the hydrophobic surface of **b'** as is observed for human PDI **b'x** (Fig 7b). Were human PDI **b'x** to adopt such an arrangement as found in yeast Pdi1p, W347 in the **x** region would be solvent exposed and highly mobile, consistent with the fluorescence and NMR data for the 'uncapped' conformer of the wild-type human **b'x** protein. We infer that the structure of the 'uncapped' conformer of isolated wild-type human **b'x** has some similarities to that observed in the crystal structure of yeast Pdi1p and that the identified 'red-shifting' mutations act by stabilising this structure in isolated human **b'x**. However, the 'uncapped' species of wild-type PDI **b'x** shows a strong tendency to dimerize and its NMR spectrum also indicates that it undergoes conformational exchange. We infer that in this state, the **b'** region shows some disorder and two **b'** domains interact (possibly via their ligand binding sites) leaving the **x** regions freely mobile and exposed to the aqueous medium.

In analyzing their structure of yeast Pdi1p, Tian et al (13) drew attention to a 'highly hydrophobic pocket' on the surface of the **b'** domain (Fig 7b) which – together with hydrophobic patches at homologous positions on the surfaces of the **a**, **b** and **a'** domains – forms a continuous hydrophobic surface. We had previously proposed that the **b'** domain provided the principal binding site on PDI for peptides and unfolded proteins, based on studies on human PDI domains and domain combinations, (12,16). Functional studies on yeast Pdi1p mutants by Tian et al. (13) confirm that this conclusion is also valid for the yeast enzyme. In Fig 7a, the residues which contribute side-chains to the 'highly hydrophobic pocket' identified by Tian et al.

(13) are highlighted and the corresponding residues in the structurally-aligned human sequence are also highlighted. Also in Fig 7a, we identify the residues of the **b'** domain of human **b'x** which form hydrophobic contacts with the sidechains of L343 and W347 of the **x**-linker. It is clear that the **x**-linker is retained in close contact with the **b'** domain by binding in the centre of a wider hydrophobic surface which corresponds to the hydrophobic pocket identified by Tian et al (13) in the yeast Pdi1p structure and which has been identified as the putative ligand binding site (Fig 7b).

Functional significance of novel structural interaction between b' domain and x linker. This work has i) demonstrated that there is an intrinsic conformational heterogeneity in isolated human PDI **b'x** between conformers which differ in the structural relationship between the **b'** domain and the **x** region, ii) determined the structure of a mutant **b'x** showing an interaction between parts of the **x** region and sites on the **b'** domain, and iii) confirmed that, in this structure, **x** is interacting with a hydrophobic site homologous to that postulated as the ligand binding site in the structure of full-length yeast Pdi1p. Hence it appears that the **x** region can adopt alternative conformations and can potentially exchange between the 'capped' conformation, in which it occludes a site on the **b'** domain, and an 'uncapped' conformation which reveals this site. It is, of course, conceivable that these phenomena are artefacts of working with isolated domain constructs rather than with full-length PDI, but there are grounds for considering that the **x** region does cycle on and off the principal substrate binding site in full-length PDI and that this is significant as part of the function of the enzyme in facilitating protein folding. The existence of such alternative conformations for the **x** region in full-length PDI would provide an explanation for why the structure of full-length mammalian PDI has been so resistant to determination by x-ray crystallography. It is significant that the structure of yeast Pdi1p was generated from a slowly-developing crystal form in which one molecule of the protein occupies the putative ligand binding site of a second, symmetry-related molecule (13); this may have the effect of trapping the protein in the 'ligand-bound' form and inhibiting conformational exchange. This suggests that the yeast structure represents a conformation in which a protein ligand is bound, displacing the **x**-linker, whereas the human **b'x** structure represents an alternative ligand-free state and that both these conformations may be relevant to the functional cycle of the enzyme. It is well-known that in the functional cycle of GroEL/ES, the hydrophobic binding site in the apical domain of GroEL that binds unfolded protein substrates alternates between being available for binding such ligands and being occluded by interaction with part of the GroES

cap (35); the interconversion between such states involves relative re-orientations of the domains of the chaperonin molecule. Such re-orientations between domains (and especially of the **x** linker relative to its neighbouring domains) may be a critical feature of the functional cycle of PDI and the process of such re-orientations in preparations of full-length PDI may generate the conformational heterogeneity which has frustrated previous structural studies

ACKNOWLEDGEMENTS

This work was supported by the Academy of Finland, Sigrid Juselius Foundation, the University of Oulu, the Biotechnology and Biological Sciences Research Council, UK (Grant BB/D017807), and the Vice-Chancellor's Fund, Warwick University. We thank Antonia Nilsson and Vicky Marlow (Warwick University) for preliminary experiments, Russell Wallis (Department of Biochemistry, Leicester University) for performing the analytical ultracentrifugation work, and Lee Byrne (University of Kent) for helpful discussions.

FOOTNOTES

[1] The abbreviations used are: NMR, nuclear magnetic resonance; PDI, protein disulphide isomerase; Pdi1p, yeast PDI (product of the S.cerevisiae *Pdi1* gene); trx, thioredoxin.

The atomic coordinates and structure factors (3bj5) have been deposited in the Protein Databank.

REFERENCES

1. Ellgaard, L. & Ruddock, L.W. (2005) *EMBO Reports* **6** 28-32
2. Freedman, R. B. (1984) *Trends Biochem.Sci* **9** 438-441
3. Creighton, T.E., Hillson, D.A. & Freedman, R.B. (1980) *J.Mol.Biol.* **142** 43-62
4. Creighton, T.E., Bagley, C.J., Cooper, L., Darby, N.J., Freedman, R.B., Kemmink, J. & Sheikh, A. (1993) *J Mol Biol* **232** 1176-1196.
5. Hawkins, H.C., Blackburn, E.C. & Freedman, R.B. (1991) *Biochem J* **275** 349-353.
6. Darby, N.J., Freedman, R.B. & Creighton, T.E. (1994) *Biochemistry* **33** 7937-7947
7. Ruoppolo, M., Freedman, R.B., Pucci, P. & Marino, G. (1996) *Biochemistry* **35** 13636-13646
8. Edman, J.C., Ellis, L., Blacher, R.W., Roth, R.A. & Rutter, W.J. (1985) *Nature* **317** 267-270
9. Kemmink, J., Darby, N.J., Dijkstra, K., Nilges, M. & Creighton, T.E. (1997) *Curr.Biol.* **7** 239-245
10. Freedman, R.B., Klappa, P. & Ruddock, L.W. (2002) *EMBO Reports* **3** 136-140
11. Alanen, H.I., Salo, K.E.H., Pekkala, M., Siekkinen, H.M., Pirneskoski, A. & Ruddock, L.W. (2003) *Antioxid. Redox. Signal.* **5** 367-374
12. Pirneskoski, A., Klappa, P., Lobell, M., Williamson, R.A., Byrne, L., Alanen, H.I., Salo, K.E.H., Kivirikko, K.I., Freedman, R.B. & Ruddock, L.W. (2004) *J.Biol.Chem.* **279** 10374-10381
13. Tian, G., Xiang, S., Noiva, R., Lennarz, W.J. & Schindelin, H. (2006) *Cell* **124** 61-73

14. Gruber, C.W., Cemazar, M., Heras, B., Martin, J.L. & Craik, D.J. (2006) *Trends.Biochem.Sci.* **31** 455-464
15. Darby, N.J., Penka, E. & Vincentelli, R. (1998) *J.Mol.Biol.* **276** 239-247
16. Klappa, P., Ruddock, L.W. Darby, N.J. & Freedman, R.B. (1998) *EMBO J.,* **17** 927-935.
17. Klappa, P., Freedman, R.B., Langenbuch, M., Robinson, G.K. & Ruddock, L.W. (2001) *Biochem.J.* **354** 553-559
18. Pollock, S., Kozlov, G., Pelletier, M.-F., Trempe, J.-F., Jansen, G., Sitnikov, D., Bergeron, J.J.M., Gehring, K., Ekiel, I., & Thomas, D.Y. (2004) *EMBO J.* **23** 1020-1029
19. Russell, S.J., Ruddock, L.W., Salo, K.E.H., Oliver, J.D., Roebuck, Q.P., Llewellyn, D.H., Roderick, H.L., Koivunen, P., Myllyharju, J. & High, S. (2004) *J.Biol.Chem.* **279** 18861-18869
20. Silvennoinen, L., Myllyharju, J., Ruoppolo, M., Orru, S., Caterino, M., Kivirikko, K.I., & Koivunen, P. (2004) *J.Biol.Chem.* **279** 13607-13615
21. Urade, R., Okudo, H., Kato, H., Moriyama, T. & Arakaki, Y. (2004) *Biochemistry* **43** 8858-8868
22. Alanen, H.I., Williamson, R.A., Howard, M.J., Hatahet, F.S., Salo, K.E.H., Kauppila, A., Kellokumpu, S. & Ruddock, L.W. (2006) *J.Biol.Chem.* **281** 33727-33737
23. Girija, U.V., Dodds, A.W., Roscher, S., Reid, K.B.M. & Wallis, R. (2007) *J.Immunol.* **179** 455-462
24. Jancarik, J. & Kim, S,H, (1991). *J.Appl.Crystallogr.* **24**, 409-411
25. Kabsch, W. (1993) *J. Appl. Crystallogr.* **26** 795-800.
26. Kursula, P. (2004) *J. Appl. Crystallogr.* **37** 347-348
27. Panjikar S., Parthasarathy V., Lamzin V.S., Weiss M.S. & Tucker P.A.(2005) *Acta Crystallogr. D* **61** 449-457.
28. Emsley, P. & Cowtan, K. (2004). *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132
29. Murshudov, A., Vagin, A. & Dodson, E.J. (1997). *Acta Crystallogr. D* **53** 240-255
30. Winn, M., Isupov, M. & Murshudov, G.N. (2001) *Acta Crystallogr. D* **57,** 122-133
31. Plaxco, K.W., Morton, C.J., Grimshaw, S.B., Jones, J.A., Pitkeathly, M., Campbell, I.D. & Dobson, C. M. (1997) *J. Biomol. NMR* **10** 221-230
32. Lambert, N. & Freedman, R.B. (1983) *Biochem.J.* **213** 225-234
33. Vuori, K., Myllyla, R., Pihlajaniemi, T. & Kivirikko, K.I. (1992) *J.Biol.Chem.* **267** 7211-7214
34. Holm, L. & Sander, C. (1996) *Science* **273**, 595-602
35. Horwich A.L., Farr, G.W. & Fenton, W.A. (2006) *Chem.Rev.* 106, 1917-1930
36. Koradi, R., Billeter, M. & Wüthrich, K. (1996) *J. Mol Graphics* **14** 51-55

FIGURE LEGENDS

Figure 1. Spectroscopic properties of human PDI **b'x**

(a) Representative relative fluorescence emission spectra, excitation at 280nm and (b) $^{15}N/^{1}H$ HSQC NMR spectra for unfractionated purified wt protein. The dashed box in (b) surrounds the two Trp indole cross-peaks.

Figure 2. Resolution of conformers of **b'x** by ion-exchange chromatography.

(a) Samples purified by IMAC were further resolved by ion-exchange chromatography as described in Methods; the solid line represents absorbance at

280 nm and the dotted line shows the conductivity. Fractions were pooled as indicated by the horizontal bars. (b and c) Re-chromatography of stored fractions; each pool was split in two and stored for a week; half was stored in the refrigerator (solid line) and half in freezer (dashed line). (b) Re-chromatography of pool 1 after storage, (c) Re-chromatography of pool 2 after storage.

Figure 3. Characterization of subfractionated **b'x** pool samples.
(a) CD spectra of 'early' (solid line) and 'late' (dashed line) pooled samples. Data are from an average of 8 scans with 10μM protein in 20mM sodium phosphate buffer, pH 7.3. (b) Denaturation of **b'x** by guanidine hydrochloride (GdmCl). 'Early' pool **b'x** is shown by the closed circles, 'late' pool **b'x** by the open circles. Fluorescence measurements were collected from 2μM protein in 20mM sodium phosphate buffer containing 150mM NaCl, pH 7.1. Samples were excited at 280nm and the data shown are a ratio of the average fluorescence between 330-334nm and 348-352nm (the λ_{max} values for the fluorescence peaks of the 'early' pool in 0M and 3.5M GdmCl respectively). (c) $^{15}N/^{1}H$ HSQC NMR spectra for the 'early' pool and (d) the 'late' pool of **b'x**.

Figure 4. Size analysis of **b'x** samples from ion-exchange chromatography.
(a) Ion exchange chromatogram (solid line, UV absorbance; dotted line, conductivity). The upper insert panel shows reducing SDS-PAGE analysis of fractions across the major peak as indicated by the horizontal bar, the lower insert panel shows native-PAGE analysis of the same column fractions. (b) Analytical gel-filtration of **b'x** pools from ion-exchange chromatography. The dashed line shows pool 1 material (fractions 1 to 4) and the solid line shows pool 2 material (fractions 6-10).

Figure 5. Screening of **b'x** mutants to identify mutations stabilising a single conformer
Representative relative fluorescence emission spectra, excitation at 280nm, of micro-scale preparations of (a) I272A and (b) L343A mutants of human PDI **b'x**. (c) Shifts in fluorescence ratios (331-337nm/352-358nm) for all 75 mutants of human PDI **b'x** relative to the wild type protein (wild type fluorescence ratio = 1.03). The data shows the mean from at least 2 independent measurements. (d) $^{15}N/^{1}H$ HSQC NMR spectra of I272A and (e) L343A mutants.

Figure 6. Crystal structure of **b'x** I272A

(a) Ribbon diagram showing the trx fold of **b'x** with the **x** region coloured in green. (b) Surface charge density for the **b'** domain in the same orientation as (a) with the **x** region shown in ribbon format. Blue = positive charge; red = negative charge; white = neutral charge. (c) Ribbon diagram with the **x** region in the front of the molecule. The side-chains of M339 (magenta), L343 (green) and W347 (green) are shown as sticks. (d) Surface charge density map for the **b'** domain in the same orientation as (c). This figure was produced using MOLMOL (36).

Figure 7. Structural comparison of human and yeast **b'x** domains

(a) Structure-based alignment of the human and yeast **b'x** sequences. A cyan background denotes residues identified as part of the hydrophobic binding site in the yeast structure (13) and the corresponding residues in the human protein. Two further residues (F287 and M307) appear to contribute to the human hydrophobic site (shown underlined in red) whereas two residues (N281 and R299) appear to fall outside the site (shown underlined in black). Residues in the human sequence coloured red comprise the binding site for W347 and L343 as judged from the **b'x** structure (at least one side-chain to side-chain atom distance within 5Å). The **x** region for both sequences is shown on a grey background with the hydrophobic residues in **x** important for interaction with **b'** shown in green and magenta. The residue numbering is for the mature protein in both cases and the secondary structure assignments (as defined in the pdb header for yeast (2b5e) and using the Kabsch and Sander algorithm as implemented in MOLMOL (36) for human) are shown above or below the sequence as appropriate. (b and c) Surface representation of the hydrophobic binding sites as defined in the sequence alignment above (those residues highlighted in cyan and underlined in red) for yeast and human **b'x** respectively. The side-chains of the hydrophobic sites are shown in white whilst the backbone and all other **b'** residues are coloured blue. The **x** regions are shown in ribbon format and the side-chains of L343 and W347 in the human structure are shown in green and M339 in magenta. The two views are in the same orientation (achieved by superposition of the core β-sheet) corresponding to that in Fig 6c and 6d. This figure was generated using MOLMOL (36).

Table 1. Statistics for the model of the crystal structure of the non-labelled human PDI **b'x** domain I272A mutant grown at pH 8.5. The values in parentheses are for the highest resolution shell.

Data collection statistics	
Space group	P3(1)21
Unit cell parameters.	
a, b, c (Å)	57.37, 57.37, 68.31
Temperature (K)	100
Wavelength (Å)	0.89997
Resolution (Å)	50-2.2
Rmerge (%)	5.5 (29.3)
Completeness (%)	99.3 (99.6)
I/σI	17.84 (5.0)
Unique reflections	6879 (849)
Redundancy	4.3 (4.4)
Mosaicity (°)	0.17
B-factor from Wilson plot ($Å^2$)	41
Refinement statistics	
Resolution (Å)	50-2.2
Total number of reflections	6877
Working set: number of reflections	6533
R_{factor} (%)	19.4
Test set: number of reflections	344
R_{free} (%)	25.5
Protein atoms (A,B,C,D)	1063
Water atoms	49
SO_4^{2-} atoms	5
Geometry statistics	
Rmsd (bond distance) (Å)	0.012
Rmsd (bond angle) (°)	1.4
Rmsd B	
Main chain bonded atoms ($Å^2$)	0.8
Side chain bonded atoms ($Å^2$)	1.8
Main chain angle ($Å^2$)	1.3
Side chain angle ($Å^2$)	2.8

Average B	
Main chain atoms ($Å^2$)	36.4
Side chain atoms ($Å^2$)	37.7
Water molecules ($Å^2$)	39.5
Sulfate atoms ($Å^2$)	80.7
Ramachandran plot	
Most favored region (%)	94
Additionally allowed regions (%)	6
Generously allowed regions (%)	0
Disallowed regions (%)	0

APPENDIX 2 - CHEMICAL SHIFT DATA FOR BB'X AT 25^{O}C.

The chemical shift data (in ppm) is presented below for amide nitrogen (^{15}N), amide proton (^{1}H) and the Cα and Cβ (^{13}C), un-assigned shifts highlighted with – and unapplicable shifts with N/A.

Residue	construct Number	Mature PDI number	^{1}HN	^{15}N	^{13}Cα	^{13}Cβ
Ala	1	119	8.4	123.9	55.8	20
Ala	2	120	7.5	117.8	50.9	20.9
Thr	3	121	8.6	119.1	70.5	62
Thr	4	122	8.7	124.9	68.6	64
Leu	5	123	8.9	126.4	49.8	41.6
Pro	6	124	-	-	-	-
Asp	7	125	7.1	112	52.4	42.1
Gly	8	126	9	106.5	47.2	N/A
Ala	9	127	8	125.1	47.3	17.4
Ala	10	128	8.2	121.7	50.7	18.2
Ala	11	129	7.6	120.5	53.8	18.5
Glu	12	130	8.4	117	55.3	30.1
Ser	13	131	7.9	112.7	61.2	62.8
Leu	14	132	7.4	122.9	58.4	41.4
Val	15	133	7.9	119.3	58.6	31.5
Glu	16	134	8.1	115	57.9	29.4
Ser	17	135	7.5	112.5	64	57.6
Ser	18	136	7	115.5	57.2	65.2
Glu	19	137	8.6	124.4	59.7	30.4
Val	20	138	7.3	113.5	60.6	35.8
Ala	21	139	9.1	126.9	51.5	22.8
Val	22	140	7.9	120.1	60.5	34.3
Ile	23	141	9	124.4	60.5	34.3
Gly	24	142	8.6	115.7	44.5	N/A
Phe	25	143	8.4	127.9	43.1	36.5
Phe	26	144	8.1	116.9	55.6	43.1
Lys	27	145	9.4	124.6	55.6	31.7
Asp	28	146	8.8	117.1	51.8	31.6
Val	29	147	8.5	121.1	63.8	30.4
Glu	30	148	8.2	117.1	55.2	29.3
Ser	31	149	7.5	115.9	58.1	65.3
Asp	32	150	8.8	121.3	58.2	40.1
Ser	33	151	8.4	113.9	57.9	62.5
Ala	34	152	6.9	125.3	54.5	18.6
Lys	35	153	8	115.9	60.1	31.8
Gln	36	154	7.8	118.9	56.3	28.5
Phe	37	155	7.7	119.9	57.4	39.2
Leu	38	156	-	-	-	-
Gln	39	157	7.6	117.7	58.6	28.5
Ala	40	158	7.6	123.6	55.2	17.2
Ala	41	159	7.2	118	53.1	17.6
Glu	42	160	6.9	112.7	57.4	30.1

Ala	43	161	7.2	119.7	52.7	20.3
Ile	44	162	7	118.9	59.8	39.4
Asp	45	163	8.3	122.9	54.5	42.2
Asp	46	164	8.5	115.7	54.7	40.3
Ile	47	165	7.1	118.9	57.6	42
Pro	48	166	-	-	-	-
Phe	49	167	8.6	120.2	42.1	32.5
Gly	50	168	9.4	109.4	43.4	N/A
Ile	51	169	8.9	120.7	58.4	43.2
Thr	52	170	8.6	121.2	72.2	60.2
Ser	53	171	8.6	119.8	56.4	66
Asn	54	172	-	-	-	-
Ser	55	173	-	-	-	-
Asp	56	174	8.6	122.6	57.4	40.1
Val	57	175	8	122.2	56.8	30.8
Phe	58	176	-	-	-	-
Ser	59	177	8	112	61.4	62.7
Lys	60	178	7.6	124	58.8	32.6
Tyr	61	179	7	113.7	57.9	37.8
Gln	62	180	7.8	113.5	57.3	30
Leu	63	181	8	120.4	54.2	40.4
Asp	64	182	8.5	120.3	53.2	41.8
Lys	65	183	7.7	117.2	54.2	34.1
Asp	66	184	7.7	116.9	54.4	41.4
Gly	67	185	8.6	107.2	45.6	N/A
Val	68	186	8.4	117.8	61.1	35.3
Val	69	187	8.9	126.8	61.3	35.9
Leu	70	188	8.6	127.3	54.3	32
Phe	71	189	9.5	126.5	56.9	42.3
Lys	72	190	8.4	116	53.5	34.6
Lys	73	191	7.9	121.3	56	32
Phe	74	192	6.6	113.4	54.4	40.4
Asp	75	193	8.8	117.6	56.2	40.4
Glu	76	194	10.1	123.6	58.3	28.7
Gly	77	195	7.7	108.9	46	N/A
Arg	78	196	7.2	121	54.9	32.6
Asn	79	197	9.2	121.6	53.5	43.8
Asn	80	198	9	123.2	52.3	43.9
Phe	81	199	8.3	124.8	58.9	39.1
Glu	82	200	7.3	128.5	54.9	32
Gly	83	201	6.9	108.4	43.4	N/A
Glu	84	202	8.3	120.5	55.7	30
Val	85	203	9	126.4	64.2	30
Thr	86	204	7.2	117.7	58	72.8
Lys	87	205	9.3	124.1	60.7	32.2
Glu	88	206	9	116.6	60.9	28.7
Asn	89	207	7.9	118.1	55.3	37.5
Leu	90	208	8.5	121.5	58.2	42.2
Leu	91	209	8.4	120	58.9	40.7
Asp	92	210	7.6	118.7	57.4	41.2
Phe	93	211	7.9	120.8	60.3	39.6
IleN	94	212	8.7	121.2	65.4	38.1
Lys	95	213	8.2	117.9	59.3	31.9
His	96	214	7.8	113.9	58.4	28.9
Asn	97	215	7.3	114.2	54.5	41

Gln	98	216	7.6	115	57.5	28.7
Leu	99	217	7.4	119.5	51.9	42.1
Pro	100	218	-	-	-	-
Leu	101	219	9.3	123.1	58	42.6
Val	102	220	7.6	109.9	59.5	34.7
Ile	103	221	8.6	124.6	60.4	41.4
Glu	104	222	8.3	127.8	54.6	29.9
Phe	105	223	8.6	129	59.2	40.2
Thr	106	224	-	-	-	-
Glu	107	225	8.9	120.2	59.2	29.1
Gln	108	226	7.9	114.6	58.1	28.8
Thr	109	227	7.4	108.8	61.9	69.5
Ala	110	228	7.6	125.5	56.7	16.1
Pro	111	229	-	-	-	-
Lys	112	230	7.2	115.5	56.7	32.3
Ile	113	231	7.6	118.9	60.9	38.7
Phe	114	232	7.9	115.4	59.3	38.4
Gly	115	233	7.6	106.7	45.2	N/A
Gly	116	234	7.3	108	44.6	N/A
Glu	117	235	8.4	117.5	57.9	30.6
Ile	118	236	7.8	119.6	62.1	37.5
Lys	119	237	8.5	125.4	55.5	33.2
Thr	120	238	6.8	115.9	62.3	70
His	121	239	8.8	125.7	53.6	35.2
Ile	122	240	9.2	122.6	58.8	35.4
Leu	123	241	9	128.6	53.8	43.7
Leu	124	242	8.6	123.3	52.9	41.8
Phe	125	243	9.1	127.5	56.9	39.9
Leu	126	244	8.9	126.8	50.8	44.5
Pro	127	245	-	-	-	-
Lys	128	246	8.2	123.7	58.9	32.4
Ser	129	247	7.6	108.2	58.5	63.9
Val	130	248	7.4	122.2	62.1	32.8
Ser	131	249	8.3	120.2	59.7	63
Asp	132	250	8.9	122.2	54.1	39.1
Tyr	133	251	7.6	120.5	63	39.5
Asp	134	252	8.6	115.3	57.4	40.4
Gly	135	253	7.8	110.1	46.7	N/A
Lys	136	254	8	123.1	59.9	33.3
Leu	137	255	8	120	57.6	40.7
Ser	138	256	8.3	114.6	57.4	62.3
Asn	139	257	8	121.7	56.7	38.8
Phe	140	258	7.7	121.7	61.4	40.6
Lys	141	259	8.4	117.6	61.4	32.6
Thr	142	260	8.2	116	59.6	66.8
Ala	143	261	7.5	123.7	54	18.8
Ala	144	262	7.3	119.6	53.7	19.7
Glu	145	263	7	114.4	58.9	30.1
Ser	146	264	7.4	111.9	60.7	63.6
Phe	147	265	7.2	115.8	57.5	41.3
Lys	148	266	7.2	123.8	58.6	31.9
Gly	149	267	9.6	115.3	45.4	N/A
Lys	150	268	8.3	118.7	57.1	35.6
Ile	151	269	7.8	116.6	60.5	43.2
Leu	152	270	8.3	129.6	54.7	43.3

Phe	153	271	9.2	129.8	56.6	39.2
Ile	154	272	9.1	123.9	59	42.1
Phe	155	273	8.9	123.1	54.6	42.4
Ile	156	274	8.7	122	59.9	42.9
Asp	157	275	8.4	123.9	60.6	40.1
Ser	158	276	7.5	124	60.3	64.9
Asp	159	277	8.4	119.6	55.7	45
His	160	278	7.2	121.3	58.4	31.5
Thr	161	279	8	121.8	62	68.9
Asp	162	280	10.7	124.3	56.7	40.1
Asn	163	281	7.9	115	52.7	38.4
Gln	164	282	7.6	121	59.3	28.1
Arg	165	283	8.4	117.9	59.4	29.1
Ile	166	284	7.7	119.6	61	40.7
Leu	167	285	8.3	119.3	57.5	40.5
Glu	168	286	7.2	119.1	59.2	29.8
Phe	169	287	8.1	121.8	60.7	39
Phe	170	288	7.5	113	59.6	39.8
Gly	171	289	8	110.7	46.7	N/A
Leu	172	290	7.7	121	54	45.9
Lys	173	291	8.1	119.1	53.9	34.3
Lys	174	292	9	124	53.9	32.1
Glu	175	293	9.4	116.6	58.7	28.1
Glu	176	294	7.6	118.1	55.8	30.5
Cys	177	295	7	117.6	58.3	27.3
Pro	178	296	-	-	-	-
Ala	179	297	8.4	123.1	51.4	23.4
Val	180	298	8.6	118.1	59.2	35.4
Arg	181	299	8.9	123.1	53.2	35.5
Leu	182	300	8.7	125.1	52.8	47.4
Ile	183	301	9.5	121.8	58.6	42.7
Thr	184	302	8.7	114	59.2	70.8
Leu	185	303	8.6	125.2	54.4	42.6
Glu	186	304	7.5	123.3	56.3	29.9
Glu	187	305	8.5	122.2	59.3	29.3
Glu	188	306	7.9	116.3	59.3	29.1
Met	189	307	8.8	123.8	56	33.8
Thr	190	308	8.4	121.7	67.5	62.3
Lys	191	309	8.6	123.4	55	36.3
Tyr	192	310	9.5	119.6		
Lys	193	311	9.2	122.4	52.6	34.8
Pro	194	312	-	-	-	-
Glu	195	313	8.9	119.5	57.5	31.5
Ser	196	314	7.5	111.7	56.4	65.4
Glu	197	315	8.7	122.9	65.4	29.3
Glu	198	316	8.2	121.2	57.4	31.2
Leu	199	317	8.6	125.4	53.1	41.3
Thr	200	318	6.8	108.3	59.5	70.5
Ala	201	319	9.2	125.3	56.1	18.1
Glu	202	320	9.3	118.2	60.3	29
Arg	203	321	7.6	119.5	58	30.1
Ile	204	322	8.5	121.8	65.4	30
Thr	205	323	8.6	116.9	68.2	60.5
Glu	206	324	8	121.7	59.8	30.5
Phe	207	325	7.9	119.9	59.3	38.7

Cys	208	326	7.7	115.9	29.9	27.7
His	209	327	8.6	118	60.1	30.2
Arg	210	328	8.7	118.9	59.9	30.2
Phe	211	329	8.3	120.6	60.2	37.3
Leu	212	330	8.2	123.6	57.4	41.5
Glu	213	331	7.7	116.3	56.3	30.6
Gly	214	332	7.7	108.5	46.5	N/A
Lys	215	333	8	116.6	56.1	33.8
Ile	216	334	7.8	118.5	59.1	38.5
Lys	217	335	8.6	128	54.5	32.4
Pro	218	336	-	-	-	-
His	219	337	8.4	120.4	57.3	29.9
Leu	220	338	7.8	118.7	53.3	41.1
Met	221	339	7.7	120.4	55.8	33
Ser	222	340	8.1	116.3	58.4	63.4
Gln	223	341	8.2	121.4	55.6	30.4
Glu	224	342	7.9	120.6	57.9	30.7
Leu	225	343	8	125.1	55.1	42.8
Pro	226	344	-	-	-	-
Glu	227	345	8.3	121.3	58	29.9
Asp	228	346	8.2	119.2	54.1	40.4
Trp	229	347	7.8	120.6	53.8	29.1
Asp	230	348	8	120.8	53.6	39.8
Lys	231	349	7.5	120.3	55.8	40.6
Gln	232	350	8.1	122.9	53.4	28.8
Pro	233	351	-	-	-	-

APPENDIX 3 - CHEMICAL SHIFT DATA FOR B'X AT 25^{O}C.

The chemical shift data (in ppm) is presented below for amide nitrogen (^{15}N), amide proton (^{1}H) and the Cα and Cβ (^{13}C), un-assigned shifts highlighted with – and unapplicable shifts with N/A.

Residue	construct Number	Mature PDI number	^{1}HN	^{15}N	^{13}Cα	^{13}Cβ
Lys	1	213	6.9	112.2	53.5	38.8
His	2	214	8.3	120.1	55.9	29.4
Asn	3	215	7.6	116.9	53.2	41.8
Gln	4	216	7.9	120.5	55.8	33.2
Leu	5	217	8.5	124.0	53.1	28.8
Pro	6	218	-	-	-	-
Leu	7	219	8.2	120.5	42.9	32.2
Val	8	220	7.4	111.6	59.9	33.9
Ile	9	221	8.7	125.4	59.8	41.7
Glu	10	222	8.4	127.6	54.7	29.9
Phe	11	223	8.7	129.4	59.3	40.7
Thr	12	224	7.1	118.2	59.1	73.2
Glu	13	225	9.1	120.6	59.6	29.1
Gln	14	226	7.9	114.5	58.3	28.8
Thr	15	227	7.5	109.0	61.8	69.6
Ala	16	228	7.6	125.4	56.8	15.9
Pro	17	229	-	-	-	-
Lys	18	230	7.2	115.5	57.6	32.3
Ile	19	231	7.6	119.5	64.3	38.5
Phe	20	232	7.9	115.2	59.7	38.4
Gly	21	233	7.7	106.8	45.3	N/A
Gly	22	234	7.4	107.8	44.8	N/A
Glu	23	235	8.5	117.3	58.2	30.6
Ile	24	236	7.8	119.6	61.8	37.8
Lys	25	237	8.6	125.5	55.4	33.0
Thr	26	238	6.8	116.0	62.1	69.8
His	27	239	8.9	125.9	54.1	-
Ile	28	240	9.3	122.5	58.9	40.2
Leu	29	241	9.1	129.0	53.6	43.4
Leu	30	242	8.7	123.3	53.0	41.5
Phe	31	243	9.2	128.1	57.0	39.0
Leu	32	244	9.1	127.2	50.8	44.4
Pro	33	245	-	-	-	-
Lys	34	246	8.3	124.1	58.9	32.2
Ser	35	247	7.6	108.2	58.4	63.7
Val	36	248	7.5	123.0	62.2	32.5
Ser	37	249	8.5	120.9	59.5	62.8
Asp	38	250	9.1	122.9	54.2	39.4
Tyr	39	251	7.6	120.2	63.1	39.4
Asp	40	252	8.7	115.0	57.4	40.6
Gly	41	253	7.8	110.2	47.1	N/A

Lys	42	254	8.0	123.3	60.0	33.2
Leu	43	255	8.2	119.8	57.5	40.4
Ser	44	256	8.4	114.7	57.6	62.6
Asn	45	257	8.0	121.9	56.3	38.7
Phe	46	258	7.7	122.0	56.4	40.4
Lys	47	259	8.5	117.2	59.8	32.5
Thr	48	260	8.3	116.6	66.5	68.7
Ala	49	261	7.7	124.6	55.0	18.7
Ala	50	262	7.3	119.1	53.9	19.9
Glu	51	263	7.0	113.8	58.8	30.0
Ser	52	264	7.5	112.0	60.8	63.5
Phe	53	265	7.2	115.9	57.1	41.3
Lys	54	266	7.4	123.4	58.9	31.8
Gly	55	267	9.6	115.1	45.6	N/A
Lys	56	268	8.4	118.9	57.3	35.7
Ile	57	269	7.9	116.4	60.4	43.3
Leu	58	270	8.4	130.6	54.6	43.1
Phe	59	271	9.3	130.4	56.5	39.1
Ile	60	272	9.1	123.7	59.0	42.2
Phe	61	273	8.9	122.2	54.7	42.5
Ile	62	274	8.8	121.1	59.9	42.8
Asp	63	275	8.5	123.5	60.0	52.6
Ser	64	276	7.5	123.9	60.5	64.7
Asp	65	277	8.3	119.0	55.7	41.5
His	66	278	7.3	121.5	58.4	31.9
Thr	67	279	8.9	127.6	65.8	69.0
Asp	68	280	11.2	125.2	56.8	39.9
Asn	69	281	8.0	115.3	52.7	38.6
Gln	70	282	7.6	120.2	59.5	28.5
Arg	71	283	7.7	118.2	55.9	30.4
Ile	72	284	8.1	120.9	64.2	37.2
Leu	73	285	7.5	122.4	60.6	41.2
Glu	74	286	8.0	117.4	59.3	29.4
Phe	75	287	8.4	122.4	58.3	29.5
Phe	76	288	7.6	112.8	60.0	39.6
Gly	77	289	8.0	110.9	46.7	N/A
Leu	78	290	7.8	121.2	53.5	46.0
Lys	79	291	8.2	119.2	53.0	34.3
Lys	80	292	9.1	124.5	61.1	32.3
Glu	81	293	9.5	116.5	58.7	28.2
Glu	82	294	7.7	117.8	53.2	30.7
Cys	83	295	7.1	117.8	58.4	27.6
Pro	84	296	-	-	-	-
Ala	85	297	8.4	123.2	51.7	23.4
Val	86	298	8.7	118.2	59.4	35.3
Arg	87	299	9.0	123.3	53.1	35.5
Leu	88	300	8.8	125.7	54.4	47.0
Ile	89	301	9.7	122.1	58.3	42.9
Thr	90	302	8.9	113.7	59.0	71.0
Leu	91	303	8.7	124.8	54.5	42.5
Glu	92	304	7.6	123.5	56.5	29.8
Glu	93	305	8.6	122.6	59.4	29.1
Glu	94	306	8.0	116.1	54.7	31.6
Met	95	307	9.1	125.2	56.5	34.2

Thr	96	308	8.4	121.3	62.1	70.8
Lys	97	309	8.8	123.6	54.4	36.3
Tyr	98	310	9.6	120.0	57.6	41.4
Lys	99	311	9.3	122.6	52.7	34.9
Pro	100	312	-	-	-	-
Glu	101	313	9.1	119.5	57.6	30.4
Ser	102	314	7.6	112.1	56.5	65.4
Glu	103	315	8.8	122.6	56.2	29.3
Glu	104	316	8.0	120.8	57.8	31.2
Leu	105	317	8.7	125.2	53.1	41.3
Thr	106	318	6.8	107.9	59.5	70.5
Ala	107	319	9.4	125.6	55.9	17.8
Glu	108	320	9.4	118.4	60.4	29.2
Arg	109	321	7.7	119.9	57.9	29.8
Ile	110	322	8.6	122.1	65.4	38.3
Thr	111	323	8.6	117.0	67.7	68.8
Glu	112	324	8.1	121.7	60.0	30.5
Phe	113	325	8.0	119.9	59.5	38.8
Cys	114	326	7.7	116.0	64.3	27.7
His	115	327	8.6	118.1	60.5	30.2
Arg	116	328	8.8	119.3	60.1	30.4
Phe	117	329	8.4	121.1	59.5	37.8
Leu	118	330	8.2	124.8	57.5	41.4
Glu	119	331	7.7	116.5	56.1	30.6
Gly	120	332	7.8	108.6	46.5	N/A`
Lys	121	333	8.0	116.6	56.3	33.8
Ile	122	334	8.0	119.4	58.7	38.2
Lys	123	335	8.9	129.3	54.7	32.7
Pro	124	336	-	-	-	-
His	125	337	8.5	120.2	55.2	31.4
Leu	126	338	8.4	126.0	55.5	42.2
Met	127	339	8.3	120.8	56.4	32.6
Ser	128	340	8.3	116.5	63.3	58.4
Gln	129	341	8.0	120.3	55.2	31.1
Glu	130	342	8.3	120.9	55.5	29.7
Leu	131	343	8.3	125.0	53.0	41.7
Pro	132	344	-	-	-	-
Glu	133	345	8.6	120.7	57.2	30.1
Asp	134	346	8.3	119.2	54.1	40.4
Trp	135	347	7.8	120.8	58.0	29.2
Asp	136	348	8.1	120.7	53.9	40.2
Lys	137	349	7.6	120.6	56.1	33.0
Gln	138	350	8.2	123.3	53.5	28.9
Pro	139	351	-	-	-	-
Val	140	352	8.4	122.2	63.4	33.0

APPENDIX 4 - H/D EXCHANGE EXPERIMENTAL DATA

b domain residues after 5 minutes in H/D exchange experiment.

residue	construct Number	Mature PDI number
Thr	3	121
Leu	5	123
Asp	7	125
Ala	10	128
Ala	11	129
Glu	12	130
Ser	13	131
Glu	16	134
Ala	21	139
Val	22	140
Ile	23	141
Gly	24	142
Phe	25	143
Phe	26	144
Lys	27	145
Glu	30	148
Ser	31	149
Ala	34	152
Lys	35	153
Gln	36	154
Phe	37	155
Leu	38	156
Gln	39	157
Ala	40	158
Ala	41	159
Glu	42	160
Ala	43	161
Ile	44	162
Ile	47	165
Phe	49	167
Gly	50	168
Ile	51	169
Thr	52	170
Asp	56	174
Phe	58	176
Lys	60	178
Tyr	61	179
Gln	62	180
Leu	63	181
Gly	67	185
Val	68	186
Val	69	187
Leu	70	188
Phe	71	189

Glu	82	200
Thr	86	204
Asn	89	207
Leu	91	209
Asp	92	210
Phe	93	211
Lys	95	213
Asn	97	215

b domain residues after 21 hours in H/D exchange experiment.

residue	construct Number	Mature PDI number
Leu	5	123
Ile	23	141
Gly	24	142
Phe	25	143
Phe	26	144
Lys	35	153
Phe	37	155
Leu	38	156
Gln	39	157
Ala	40	158
Ala	41	159
Glu	42	160
Gly	50	168
Ile	51	169
Thr	52	170
Phe	58	176
Gly	67	185
Val	68	186
Leu	70	188
Phe	71	189
Asn	89	207
Leu	91	209

bb'x domain residues after 5 minutes in H/D exchange experiment.

residue	construct Number	Mature PDI number
Leu	5	123
Asp	7	125
Glu	12	130
Ser	13	131
Leu	14	132
Val	15	133
Glu	16	134
Ser	17	135
Glu	19	137
Val	20	138

Ala	21	139
Ile	23	141
Gly	24	142
Phe	25	143
Phe	26	144
Glu	30	148
Ser	31	149
Ala	34	152
Lys	35	153
Gln	36	154
Gln	39	157
Ala	41	159
Glu	42	160
Ile	44	162
Phe	49	167
Gly	50	168
Ile	51	169
Thr	52	170
Asn	54	172
Lys	60	178
Tyr	61	179
Gln	62	180
Gly	67	185
Val	68	186
Val	69	187
Phe	71	189
Lys	72	190
Phe	81	199
Glu	82	200
Thr	86	204
Leu	90	208
Leu	91	209
Asp	92	210
Phe	93	211
IleN	94	212
Lys	95	213
Val	102	220
Phe	105	223
Lys	112	230
Ile	113	231
Ile	122	240
Leu	123	241
Leu	124	242
Phe	125	243
Leu	126	244
Lys	128	246
Phe	140	258
Thr	142	260
Ala	143	261
Ala	144	262
Phe	147	265
Ile	151	269
Leu	152	270
Ile	154	272
Phe	155	273

Ile	156	274
Ser	158	276
Thr	161	279
Ile	166	284
Leu	167	285
Leu	172	290
Glu	176	294
Cys	177	295
Ala	179	297
Val	180	298
Arg	181	299
Ile	183	301
Glu	197	315
Thr	200	318
Ile	204	322
Thr	205	323
Phe	207	325
Cys	208	326
Arg	210	328
Glu	213	331
Lys	215	333
Glu	227	345
Asp	228	346
Trp	229	347
Asp	230	348
Lys	231	349

bb'x domain residues after 21 hours in H/D exchange experiment.

residue	construct Number	Mature PDI number
Leu	5	123
Val	15	133
Ala	21	139
Ile	23	141
Gly	24	142
Phe	25	143
Phe	26	144
Gln	39	157
Ala	43	161
Gly	50	168
Ile	51	169
Thr	52	170
Lys	60	178
Gly	67	185
Phe	71	189
Glu	82	200
Leu	91	209
IleN	94	212
Ile	122	240
Leu	123	241
Leu	124	242
Phe	125	243
Leu	126	244

Ala	143	261
Ala	144	262
Ile	151	269
Ile	154	272
Phe	155	273
Ile	156	274
Ile	166	284
Leu	167	285
Arg	181	299
Phe	207	325
Lys	217	335

APPENDIX 5 – NUCLEOTIDE AND PROTEIN SEQUENCES

b domain:

ATGCATCACCATCACCACCATNTGGNTGCCACCACCCTGCCTGACGGC
GCAGCTGCAGAGTCCTTGGTGGAGTCCAGCGAGGTGGCTGTCATCGG
CTTCTTCAAGGACGTGGAGTCGGACTCTGCCAAGCAGTTTTTGCAGGC
AGCAGAGGCCATCGATGACATACCATTTGGGATCACTTCCAACAGTGAC
GTGTTCTCCAAATACCAGCTCGACAAAGATGGGGTTGTCCTCTTTAAGA
AGTTTGATGAAGGNCGGAACAACTTTGAAGGGGAGGTCACCAAGGAGA
ACCTGNTGGACTTTATCAAACACAACCAGCTGCCCCTTGTCATCGAGTT
CACCGAGCAGACAGCC

M H H H H H H M A A T T L P D G A A A E S L V E S S E V A V I G F F K D
V E S D S A K Q F L Q A A E A I D D I P F G I T S N S D V F S K Y Q L D K
D G V V L F K K F D E G R N N F E G E V T K E N L L D F I K H N Q L P L
V I E F T E Q T A

b'x:

ATGCATCACCATCACCACCATATGAAACACAACCAGCTGCCCCTTGTCA
TTGAGTTCACCGAGCAGACAGCCCCGAAGATTTTTGGAGGTGAAATCAA
GACTCACATCCTGCTGTTCTTGCCCAAGAGTGTGTCTGACTATGACGGC
AAACTGAGCAACTTCAAAACAGCAGCCGAGAGCTTCAAGGGCAAGATC
CTGTTCATCTTCATCGACAGCGACCACACCGACAACCAGCGCATCCTC
GAGTTCTTTGGCCTGAAGAAGGAAGAGTGCCCGGCCGTGCGCCTCATC
ACCCTGGAGGAGGAGATGACCAAGTACAAGCCCGAATCGGAGGAGCT
GACGGCAGAGAGGATCACAGAGTTCTGCCACCGCTTCCTGGAGGGCAA
AATCAAGCCCCACCTGATGAGCCAGGAGCTGCCGGAGGACTGGGACA
AGCAGCCT

M H H H H H H M K H N Q L P L V I E F T E Q T A P K I F G G E I K T H I L
L F L P K S V S D Y D G K L S N F K T A A E S F K G K I L F I F I D S D H
T D N Q R I L E F F G L K K E E C P A V R L I T L E E E M T K Y K P E S E
E L T A E R I T E F C H R F L E G K I K P H L M S Q E L P E D W D K Q P

b':

ATGCATCACCATCACCACCATATGAAACACAACCAGCTGCCCCTTGTCA
TTGAGTTCACCGAGCAGACAGCCCCGAAGATTTTTGGAGGTGAAATCAA
GACTCACATCCTGCTGTTCTTGCCCAAGAGTGTGTCTGACTATGACGGC
AAACTGAGCAACTTCAAAACAGCAGCCGAGAGCTTCAAGGGCAAGATC
CTGTTCATCTTCATCGACAGCGACCACACCGACAACCAGCGCATCCTC
GAGTTCTTTGGCCTGAAGAAGGAAGAGTGCCCGGCCGTGCGCCTCATC
ACCCTGGAGGAGGAGATGACCAAGTACAAGCCCGAATCGGAGGAGCT
GACGGCAGAGAGGATCACAGAGTTCTGCCACCGCTTCCTGGAGGGC

MHHHHHHMKHNQLPLVIEFTEQTAPKIFGGEIKTHIL
LFLPKSVSDYDGKLSNFKTAAESFKGKILFIFIDSDH
TDNQRILEFFGLKKEECPAVRLITLEEEMTKYKPESE
ELTAERITEFCHRFLEG

bb':

ATGCATCACCATCACCACCATATGGCTGCCACCACCCTGCCTGACGGC
GCAGCTGCAGAGTCCTTGGTGGAGTCCAGCGAGGTGGCTGTCATCGG
CTTCTTCAAGGACGTGGAGTCGGACTCTGCCAAGCAGTTTTTGCAGGC
AGCAGAGGCCATCGATGACATACCATTTGGGATCACTTCCAACAGTGAC
GTGTTCTCCAAATACCAGCTCGACAAAGATGGGGTTGTCCTCTTTAAGA
AGTTTGATGAAGGCCGGAACAACTTTGAAGGGGAGGTCACCAAGGAGA
ACCTGCTGGACTTTATCAAACACAACCAGCTGCCCCTTGTCATTGAGTT
CACCGAGCAGACAGCCCCGAAGATTTTTGGAGGTGAAATCAAGACTCA
CATCCTGCTGTTCTTGCCCAAGAGTGTGTCTGACTATGACGGCAAACTG
AGCAACTTCAAAACAGCAGCCGAGAGCTTCAAGGGCAAGATCCTGTTC
ATCTTCATCGACAGCGACCACACCGACAACCAGCGCATCCTGGAGTTC
TTTGGCCTGAAGAAGGAAGAGTGCCCGGCCGTGCGCCTCATCACCCTG
GAGGAGGAGATGACCAAGTACAAGCCCGAATCGGAGGAGCTGACGGC
AGAGAGGATCACAGAGTTCTGCCACCGCTTCCTGGAGGGC

MHHHHHHMAATTLPDGAAAESLVESSEVAVIGFFKD
VESDSAKQFLQAAEAIDDIPFGITSNSDVFSKYQLDK
DGVVLFKKFDEGRNNFEGEVTKENLLDFIKHNQLPL

VIEFTEQTAPKIFGGEIKTHILLFLPKSVSDYDGKLSN
FKTAAESFKGKILFIFIDSDHTDNQRILEFFGLKKEEC
PAVRLITLEEEMTKYKPESEELTAERITEFCHRFLEG

bb'x:

ATGCATCACCACCACCACCATATGGCTGCCACCACCCTGCCTGACGGC
GCAGCTGCAGAGTCCTTGGTGGAGTCCAGCGAGGTGGCTGTCATCGG
CTTCTTCAAGGACGTGGAGTCGGACTCTGCCAAGCAGTTTTTGCAGGC
AGCAGAGGCCATCGATGACATACCATTTGGGATCACTTCCAACAGTGAC
GTGTTCTCCAAATACCAGCTCGACAAAGATGGGGTTGTCCTCTTTAAGA
AGTTTGATGAAGGCCGGAACAACTTTGAAGGGGAGGTCACCAAGGAGA
ACCTGCTGGACTTTATCAAACACAACCAGCTGCCCCTTGTCATTGAGTT
CACCGAGCAGACAGCCCCNAAGATTTTTGGAGGTGAAATCAAGACTCA
CATCCTGNTGTTCTTGCCCAAGAGTGTGTCTGACTATGACGGCAAACTG
AGCAACTTCAAAACAGCAGCCGAGAGCTTCAAGGGCAAGATCCTGTTC
ATCTTCATCGACAGCGACCACACCGACAACCAGCGCATCCTGGAGTTC
TTTGGCCTGAAGAAGGAAGAGTGCCCGGCCGTGCGCCTCATCACCCTG
GAGGAGGAGATGACCAAGTACAAGCCCGAATCGGAGGAGCTGACGGC
AGAGAGGATCACAGAGTTCTGCCACCGCTTCCTGGAGGGCAAAATCAA
GCCCCACCTGATGAGCCAGGAGCTGCCGGAGGACTGGGACAAGCAGC
CT

MHHPHHHMAATTLPDGAAAESLVESSEVAVIGFFKD
VESDSAKQFLQAAEAIDDIPFGITSNSDVFSKYQLDK
DGVVLFKKFDEGRNNFEGEVTKENLLDFIKHNQLPL
VIEFTEQTAPKIFGGEIKTHILLFLPKSVSDYDGKLSN
FKTAAESFKGKILFIFIDSDHTDNQRILEFFGLKKEEC
PAVRLITLEEEMTKYKPESEELTAERITEFCHRFLEG
KIKPHLMSQELPEDWDKQP

Full length PDI:

ATGCATCACCATCACCACCATATGGACGCCCCCGAGGAGGAGGACCAC
GTCCTGGTGCTGAGGAAAAGCAACTTCGCGGAGGCGCTGGCGGCCCA
CAAGTACCTGCTGGTGGAGTTCTATGCCCCTTGGTGTGGCCACTGCAA
GGCTCTGGCCCCTGAGTATGCCAAAGCCGCTGGGAAGCTGAAGGCAG
AAGGTTCCGAGATCAGGTTGGCCAAGGTGGACGCCACGGAGGAGTCT
GACCTGGCCCAGCAGTACGGCGTGCGCGGCTATCCCACCATCAAGTTC
TTCAGGAATGGAGACACGGCTTCCCCCAAGGAATATACAGCTGGCAGA
NAGGCTGATGACATCGTGAACTGGCTGAAGAAGCGCACGGGCCCGGC
TGCCACCACCCTGCCTGACGGCGCAGCTGCAGAGTCCTTGGTGGAGTC
CAGCGAGGTGGCTGTCATCGGCTTCTTCAAGGACGTGGAGTCGGACTC
TGCCAAGCAGTTTTTGCAGGCAGCAGAGGCCATCGATGACATACCATTT
GGGATCACTTCCAACAGTGACGTGTTCTCCAAATACCAGCTCGACAAAG
ATGGGGTTGTCCTCTTTAAGAAGTTTGATGAAGGCCGGAACAACTTTGA
AGGGGAGGTCACCAAGGAGAACCTGCTGGACTTTATCAAACACAACCA
GCTGCCCCTTGTCATTGAGTTCACCGAGCAGACAGCCCCGAAGATTTTT
GGAGGTGAAATCAAGACTCACATCCTGCTGTTCTTGCCCAAGAGTGTGT
CTGACTATGACGGCAAACTGAGCAACTTCAAAACAGCAGCCGAGAGCT
TCNAGGGCAAGATCCTGTTCATCTTCATCGACAGCGACCACCCCGACA
ACCAGCGCATCCTCGAGTTTTTTNGCCTGAAGAANGAAGAGTGCCCGG
CCGTGCGCCTCATCACCCTGGAGGAGGAGATGACCAAGTACAAGCCCG
AATCGGAGGAGCTGACGGCAGAGAGGATCACAGAGTTCTGCCACCGCT
TCCTGGAGGGCAAAATCAAGCCCCACCTGATGAGCCAGGAGCTGCCG
GAGGACTGGGACAAGCAGCCTGTCAAGGTGCTTGTTGGGAAGAACTTT
GAAGACGTGGCTTTTGATGAGAAAAAAAACGTCTTTGTGGAGTTCTATG
CCCCATGGTGTGGTCACTGCAAACAGTTGGCTCCCATTTGGGATAAACT
GGGAGAGACGTACAAGGACCATGAGAACATCGTCATCGCCAAGATGGA
CTCGACTGCCAACGAGGTGGAGGCCGTCAAAGTGCACAGCTTCCCCAC
ACTCAAGTTCTTTCCTGCCAGTGCCGACAGGACGGTCATCGATTACAAC
GGGGAACGCACGCTGGATGGTTTTAAGAAATTCCTGGAGAGCGGTGGC
CAGGATGGGGCAGGGGATGATGACGATCTCGAGGACCTGGAAGAAGC
AGAGGAGCCAGACATGGAGGAAGACGATGATCAGAAAGCTGTGAAAGA
TGAACTG

MHHHHHHMDAPEEEDHVLVLRKSNFAEALAAHKYLL
VEFYAPWCGHCKALAPEYAKAAGKLKAEGSEIRLAK
VDATEESDLAQQYGVRGYPTIKFFRNGDTASPKEYT
AGREADDIVNWLKKRTGPAATTLPDGAAAESLVESS
EVAVIGFFKDVESDSAKQFLQAAEAIDDIPFGITSNS
DVFSKYQLDKDGVVLFKKFDEGRNNFEGEVTKENLL
DFIKHNQLPLVIEFTEQTAPKIFGGEIKTHILLFLPKS
VSDYDGKLSNFKTAAESFKGKILFIFIDSDHPDNQRI
LEFFGLKKEECPAVRLITLEEEMTKYKPESEELTAER
ITEFCHRFLEGKIKPHLMSQELPEDWDKQPVKVLVG
KNFEDVAFDEKKNVFVEFYAPWCGHCKQLAPIWDKL
GETYKDHENIVIAKMDSTANEVEAVKVHSFPTLKFFP
ASADRTVIDYNGERTLDGFKKFLESGGQDGAGDDD
DLEDLEEAEEPDMEEDDDQKAVKDEL